PB88-190509

International Sulfur
Deposition Model Evaluation
Final Report, May 1987

(U.S.) Environmental Protection Agency
Research Triangle Park, NC

May 87

U.S. Department of Commerce
National Technical Information Service
NTIS

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA/600/3-87/008	2.	3. RECIPIENT'S ACCESSION NO. PB88 190509/AS
4. TITLE AND SUBTITLE INTERNATIONAL SULFUR DEPOSITION MODEL EVALUATION		5. REPORT DATE May 1987 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) 1.Terry L. Clark and Robin L. Dennis 2Steve K. Seilkop,3Mayer Alvo, 4Marvin P. Olson and Eva C. Voldner		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS 1 see below 2 Analytical Sciences, Inc. Research Triangle Park, NC 3 University of Ottawa Ottawa, Ontario 4 Atmos. Envir. Service Downsview, Ontario		10. PROGRAM ELEMENT NO. N104/C/C/05 -3166 (FY-87) 11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS Atmospheric Sciences Research Laboratory - RTP, NC Office of Research and Development U.S. Environmental Protection Agency Research Triangle Park, North Carolina 27711		13. TYPE OF REPORT AND PERIOD COVERED Final 14. SPONSORING AGENCY CODE EPA/600/09
15. SUPPLEMENTARY NOTES		

16. ABSTRACT

The International Sulfur Deposition Model Evaluation (ISDME) project, jointly conducted by the U.S. Environmental Protection Agency and Atmospheric Environment Service of Environment Canada, assessed the performance of eleven linear chemistry atmospheric models in predicting amounts of sulfur wet deposition. Standardized model input data sets were distributed to the participating modelers, who later submitted seasonal and annual 1980 model predictions of dry/wet deposition and air concentrations of sulfur dioxide and sulfate at up to 65 sites across eastern North America. The models were evaluated in an operational mode using new, more rigorous approaches, as well as the more conventional distribution statistics recommended by the American Meteorological Society. The new approaches focused on the ability of the models to replicate features of the spatial patterns of sulfur wet deposition, as determined by an interpolation technique known as kriging. This technique quantified the uncertainties in the observations which were used in the evaluation process to identify areas where interpolated predictions were statistically significantly different from the interpolated observations. To supplement the evaluation, predictions of dry deposition amounts and air concentrations of each model were intercompared to identify apparent peculiarities. Finally, a scoring system based on criteria for six model performance measures was devised to compare seasonal, annual and overall performances of the models. Three clusters of models, each with similar overall scores, were identified.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

18. DISTRIBUTION STATEMENT RELEASE TO PUBLIC	19. SECURITY CLASS *(This Report)* UNCLASSIFIED	21. NO. OF PAGES 267
	20. SECURITY CLASS *(This page)* UNCLASSIFIED	22. PRICE

EPA Form 2220-1 (Rev. 4-77) PREVIOUS EDITION IS OBSOLETE

ERRATA

International Sulfur Deposition Model Evaluation

Page No.	Modification
xviii	The name of Richard Aylesworth, Environment Canada, was regretfully omitted from the Acknowledgments .
36	In the third paragraph Figure 3.4 should be Figure 3.8 while Figures 3.5-3.8 should be Figures 3.4-3.7.
39	In Table 3.4 the number of sites in overall data quality level 2 for winter should be replaced by 14.
43	In the first paragraph Figure 3.4 should be Figure 3.8.
76	In the fourth paragraph the reference to Table D.16 should be deleted.
161	The first sentence of Section 7.3 should include a reference to Appendix E.

3/1/88

PB88-190509

EPA

United States
Environmental Protection
Agency

Office of Research and
Development
Research Triangle Park NC 27711

EPA 600 3-87 008
May 1987

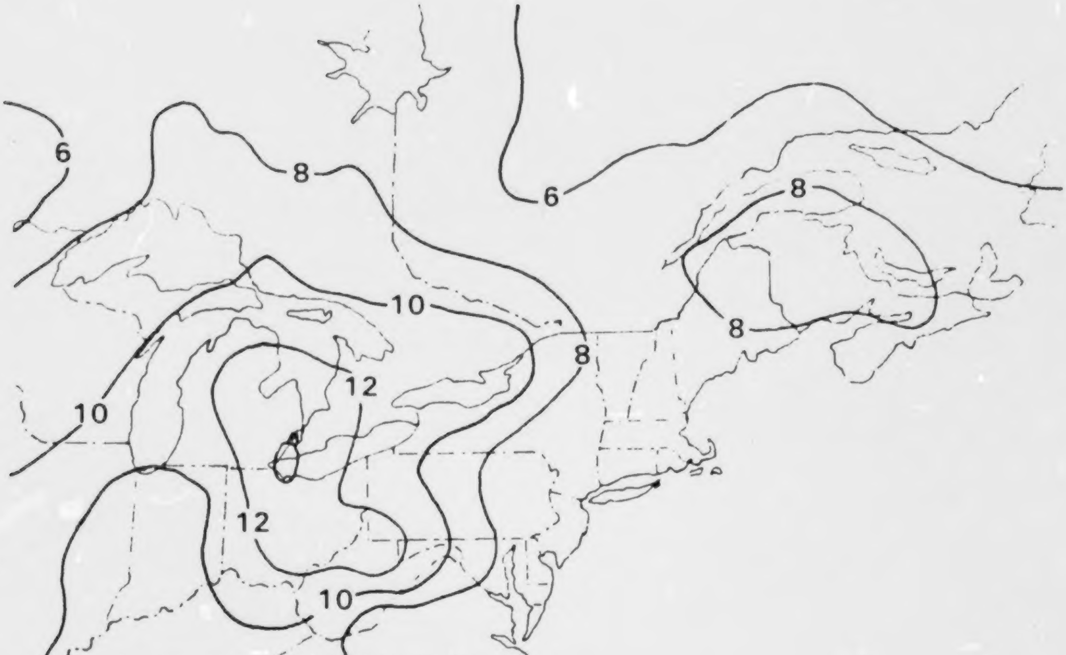
Environment
Canada

Environnement
Canada

Atmospheric Environment
Service
Downsview, Ontario M3H 5T4

Service de l'environnement
atmosphérique
Downsview, Ontario M3H 5T4

International Sulfur Deposition Model Evaluation



REPRODUCED BY
U.S. DEPARTMENT OF COMMERCE
National Technical Information Service
SPRINGFIELD, VA. 22161

EPA Publication No. EPA/600/3-87/008
AES Publication No. ARD-87-1

INTERNATIONAL SULFUR DEPOSITION MODEL EVALUATION

FINAL REPORT

May 1987

Terry L. Clark[*] and Robin L. Dennis[*]
Office of Research and Development
U.S. Environmental Protection Agency
Research Triangle Park, North Carolina

Eva C. Voldner and Marvin P. Olson
Atmospheric Environment Service
Environment Canada
Downsview, Ontario

Steve K. Seilkop
Analytical Sciences, Inc.
Research Triangle Park, North Carolina

Mayer Alvo[**]
Department of Mathematics
University of Ottawa
Ottawa, Ontario

---

[*] Also affiliated with the U.S. Department of Commerce, National Oceanic and Atmospheric Administration.

[**] Also affiliated with the Systems and Informatics Directorate, Environment Canada.

i.b

Notice

The research described in this document has been funded in part by the United States Environmental Protection Agency. It has been subject to the Agency's peer and administration review and has been approved for publication as an EPA document.

Cover Illustration

The kriged analysis of 1980 annual amounts of sulfur wet deposition (kg S/ha/yr).

ABSTRACT

The International Sulfur Deposition Model Evaluation (ISDME) project, jointly conducted by the U.S. Environmental Protection Agency and Atmospheric Environment Service of Environment Canada, assessed the performance of eleven linear chemistry atmospheric models in predicting amounts of sulfur wet deposition. Standardized model input data sets were distributed to the participating modelers, who later submitted seasonal and annual 1980 model predictions of dry/wet deposition and air concentrations of sulfur dioxide and sulfate at up to 65 sites across eastern North America.

The models were evaluated in an operational mode using new, more rigorous approaches, as well as the more conventional distribution statistics recommended by the American Meteorological Society. The new approaches focused on the ability of the models to replicate features of the spatial patterns of sulfur wet deposition, as determined by an interpolation technique known as kriging. This technique quantified the uncertainties in the observations which were used in the evaluation process to identify areas where interpolated predictions were statistically significantly different from the interpolated observations. To supplement the evaluation, predictions of dry deposition amounts and air concentrations of each model were intercompared to identify apparent peculiarities.

Finally, a scoring system based on criteria for six model performance measures was devised to compare seasonal, annual and overall performances of the models. Three clusters of models, each with similar overall scores, were identified.

CONTENTS

Page No.

ABSTRACT iii
FIGURES ix
TABLES xv
ACKNOWLEDGMENTS xviii
EXECUTIVE SUMMARY ixx

1. INTRODUCTION 1

  1.1 Background 1

  1.2 ISDME Goals 2

  1.3 ISDME Participants 4

  1.4 Structure of the Report 6

2. MODEL DESCRIPTIONS 7

3. ISDME DATA SETS 21

  3.1 Standardized Model Input Data 21

    3.1.1 General Considerations 21

    3.1.2 Transport Winds 22

    3.1.3 Precipitation Amounts 22

    3.1.4 Climatological Mixing Heights 25

    3.1.5 Anthropogenic Sulfur Dioxide Emissions Data 25

      3.1.5.1 Canadian Data 26
      3.1.5.2 United States Data 27
      3.1.5.3 Aggregation Methods 28

  3.2 Sulfur Wet Deposition Data 32

    3.2.1 Data Screening and Calculation Procedures 32

    3.2.2 Results of Data Screening and Calculations 35

    3.2.3 Differences in Amounts at Nearby Sites 40

  3.3 Air Concentration Data 44

Page No.

4. MODEL EVALUATION APPROACH 47

4.1 Introduction 47

4.2 Elements of the Model Evaluation 48

4.3 Model Performance Measures 50

4.3.1 Basis for Selection 50

4.3.2 Description of Measures 53

4.3.2.1 Spatial Patterns 53
4.3.2.2 Maxima 58
4.3.2.3 Temporal Patterns 60

4.4 Model Clustering 60

5. SPATIAL INTERPOLATION 65

5.1 Introduction 65

5.2 Selection of an Interpolation Method 66

5.3 The Use of Kriging in the Model Evaluation 70

6. MODEL EVALUATION RESULTS 75

6.1 Introduction 75

6.2 Seasonal Temporal Patterns 76

6.3 Seasonal Model Evaluation Results 78

6.3.1 Summer 1980 Results 78

6.3.1.1 Summary of Distributional Statistics 78
6.3.1.2 Comparison of Predictions and APN Measurements 82
6.3.1.3 Spatial Analyses 82
6.3.1.4 Model Performance Measures 92

Page No.

6.3.2 Spring 1980 Results 98

6.3.2.1 Summary of Distributional Statistics 98
6.3.2.2 Comparison of Predictions and APN Measurements 98
6.3.2.3 Spatial Analyses 102
6.3.2.4 Model Performance Measures 102

6.3.3 Autumn 1980 Results 114

6.3.3.1 Summary of Distributional Statistics 114
6.3.3.2 Comparison of Predictions and APN Measurements 117
6.3.3.3 Spatial Analyses 117
6.3.3.4 Model Performance Measures 128

6.3.4 Winter 1980 Results 133

6.3.4.1 Summary of Distributional Statistics 133
6.3.4.2 Comparison of Predictions and APN Measurements 133
6.3.4.3 Spatial Analyses 137
6.3.4.4 Model Performance Measures 145

6.4 Consistencies of Seasonal and Annual Results 151

7. MODEL CLUSTERING 153

7.1 Introduction 153

7.2 Overall Model Performance 154

7.2.1 Group A Models 155

7.2.2 Group B Models 158

7.2.2 Group C Models 159

7.3 Annual Model Performance 161

8. SUMMARY AND CONCLUSIONS 163

8.1 The ISDME Approach 163

8.2 Comparison of Predictions and Observations 164

Page No.

8.3 Model Performance Scores 166

8.4 Concluding Remarks 169

9. RECOMMENDATIONS FOR FUTURE RESEARCH 171

REFERENCES

APPENDICES

A SUMMARY OF PREVIOUS EVALUATIONS OF SULFUR DEPOSITION MODELS A-1

B STANDARDIZED MODEL INPUT DATA (FIGURES AND TABLES) B-1

C SCREENING AND CALCULATION PROCEDURES FOR THE MODEL EVALUATION DATA SET C-1

D COMPARISON OF PREDICTIONS AND OBSERVATIONS OF SULFUR WET DEPOSITION AT THE ISDME SITES D-1

E ANNUAL 1980 MODEL PERFORMANCES E-1

LIST OF FIGURES

Figure No.	Title	Page No.
3.1	The three layers into which sulfur dioxide emissions from point and area sources are injected.	29
3.2	The 1980 annual ISDME grids of sulfur dioxide emissions (tonnes/year) for the UMACID configuration from area sources (a) and point sources in the lowest (b), middle (c), and highest (d) layers.	30
3.3	Locations and levels of regional representativeness of the precipitation chemistry sites operating in 1980 and passing at least the Level 3 criteria.	35
3.4	**Winter 1980** amounts of sulfur wet deposition (kg S/ha) at the ISDME evaluation sites.	37
3.5	**Spring 1980** amounts of sulfur wet deposition (kg S/ha) at the ISDME evaluation sites.	37
3.6	**Summer 1980** amounts of sulfur wet deposition (kg S/ha) at the ISDME evaluation sites.	38
3.7	**Autumn 1980** amounts of sulfur wet deposition (kg S/ha) at the ISDME evaluation sites.	38
3.8	**Annual 1980** amounts of sulfur wet deposition (kg S/ha) at the ISDME evaluation sites.	39
3.9	The location and identification codes of the ISDME sites.	40
4.1	The four subregions where significant differences between interpolated predictions and observations are identified.	53
4.2	The ISDME sites in the high deposition region as denoted by the circles.	55
5.1	Comparison of the interpolated patterns of annual 1980 amounts of sulfur wet deposition (kg S/ha/yr) using a kriging technique and a standard least squares regression technique.	67
5.2	The uncertainty estimates (%) of the kriged values of annual 1980 observations of sulfur wet deposition.	71

LIST OF FIGURES

Figure No.	Title	Page No.
5.3	The histograms of kriged residuals for the summer and annual 1980 amounts of sulfur wet deposition (kg S/ha).	73
6.1	The mean percentage of annual 1980 sulfur wet deposition observed and predicted for each season at 30 ISDME sites.	77
6.2	**Summer 1980** box plots of the predictions and observations of sulfur wet deposition (kg S/ha).	79
6.3	**Summer 1980** box plots of the predictions of sulfate air concentrations ($\mu g/m^3$).	79
6.4	**Summer 1980** box plots of the predictions of sulfur dioxide air concentrations ($\mu g/m^3$).	80
6.5	**Summer 1980** box plots of the predictions of sulfur dry deposition (kg S/ha).	80
6.6	Comparison of the **summer 1980** observations and predictions of the air concentrations ($\mu g/m^3$) of sulfur dioxide and sulfate at the APN sites.	83
6.7	**Summer 1980** kriged analyses of the predictions and observations of sulfur wet deposition (kg S/ha).	84
6.8	Areas of significant under/overpredictions of **summer 1980** amounts of sulfur wet deposition.	89
6.9	**Summer 1980** locations of the centroids.	95
6.10	**Summer 1980** locations and magnitudes (kg S/ha) of the maximum amount of sulfur wet deposition at the ISDME sites.	97
6.11	**Spring 1980** box plots of the predictions and observations of sulfur wet deposition (kg S/ha).	99
6.12	**Spring 1980** box plots of the predictions of sulfate air concentrations ($\mu g/m^3$).	99
6.13	**Spring 1980** box plots of the predictions of sulfur dioxide air concentrations ($\mu g/m^3$).	100

LIST OF FIGURES

Figure No.	Title	Page No.
6.14	**Spring 1980** box plots of the predictions of sulfur dry deposition (kg S/ha).	100
6.15	Comparison of the **spring 1980** observations and predictions of the air concentrations (μg/m^3) of sulfur dioxide and sulfate at the APN sites.	101
6.16	**Spring 1980** kriged analyses of the predictions and observations of sulfur wet deposition (kg S/ha).	103
6.17	Areas of significant under/overpredictions of **spring 1980** amounts of sulfur wet deposition.	107
6.18	**Spring 1980** locations of the centroids.	112
6.19	**Spring 1980** locations and magnitudes (kg S/ha) of the maximum amount of sulfur wet deposition at the ISDME sites.	113
6.20	**Autumn 1980** box plots of the predictions and observations of sulfur wet deposition (kg S/ha).	115
6.21	**Autumn 1980** box plots of the predictions of sulfate air concentrations (μg/m^3).	115
6.22	**Autumn 1980** box plots of the predictions of sulfur dioxide air concentrations (μg/m^3).	116
6.23	**Autumn 1980** box plots of the predictions of sulfur dry deposition (kg S/ha).	116
6.24	Comparison of the **autumn 1980** observations and predictions of the air concentrations (μg/m^3) of sulfur dioxide and sulfate at the APN sites.	118
6.25	**Autumn 1980** kriged analyses of the predictions and observations of sulfur wet deposition (kg S/ha).	119
6.26	Areas of significant under/overpredictions of **autumn 1980** amounts of sulfur wet deposition.	125
6.27	**Autumn 1980** locations of the centroids.	131

LIST OF FIGURES

Figure No.	Title	Page No.
6.28	**Autumn 1980** locations and magnitudes (kg S/ha) of the maximum amount of sulfur wet deposition at the ISDME sites.	132
6.29	**Winter 1980** box plots of the predictions and observations of sulfur wet deposition (kg S/ha).	134
6.30	**Winter 1980** box plots of the predictions of sulfate air concentrations (μg/m^3).	134
6.31	**Winter 1980** box plots of the predictions of sulfur dioxide air concentrations (μg/m^3).	135
6.32	**Winter 1980** box plots of the predictions of sulfur dry deposition (kg S/ha).	135
6.33	Comparison of the **winter 1980** observations and predictions of the air concentrations (μg/m^3) of sulfur dioxide and sulfate at the APN sites.	136
6.34	**Winter 1980** kriged analyses of the predictions and observations of sulfur wet deposition (kg S/ha).	138
6.35	Areas of significant under/overpredictions of the **winter 1980** amounts of sulfur wet deposition.	142
6.36	**Winter 1980** locations of the centroids.	148
6.37	Seasonal variation of the orientation of the major axes of the annual 1980 predictions and observations of sulfur wet deposition.	149
6.38	**Winter 1980** locations and magnitudes (kg S/ha) of the maximum amount of sulfur wet deposition at the ISDME sites.	150
7.1	The distribution of the seasonal, overall and annual 1980 model performance scores.	157
B.1	The four model grid configurations for which model input data have been processed.	B-2
B.2	**Annual 1980** patterns of precipitation amounts (cm) using a) CMC gridded data and b) data from precipitation chemistry monitoring sites.	B-3

LIST OF FIGURES

Figure No.	Title	Page No.
B.3	**Winter 1980** patterns of precipitation amounts (cm) using a) CMC gridded data and b) data from precipitation chemistry monitoring sites.	B-4
B.4	**Spring 1980** patterns of precipitation amounts (cm) using a) CMC gridded data and b) data from precipitation chemistry monitoring sites.	B-5
B.5	**Summer 1980** patterns of precipitation amounts (cm) using a) CMC gridded data and b) data from precipitation chemistry monitoring sites.	B-6
B.6	**Autumn 1980** patterns of precipitation amounts (cm) using a) CMC gridded data and b) data from precipitation chemistry monitoring sites.	B-7
B.7	Histograms of 1980 annual precipitation amounts (cm) a) for 127-km, CMC grid cells containing precipitation chemistry monitoring sites and b) at the 72 precipitation chemistry monitoring sites across eastern North America.	B-8
B.8	A comparison of the weekly time series of precipitation amounts (mm) measured at 18 precipitation chemistry monitoring sites (validation data) and calculated for respective 127-km grid cells by the CMC model.	B-9
E.1	**Annual 1980** box plots of the predictions and observations of sulfur wet deposition (kg S/ha).	E-2
E.2	**Annual 1980** box plots of the predictions of sulfate air concentrations (μg/m^3).	E-2
E.3	**Annual 1980** box plots of the predictions of sulfur dioxide air concentrations (μg/m^3).	E-3
E.4	**Annual 1980** box plots of the predictions of sulfur dry deposition (kg S/ha).	E-3
E.5	**Annual 1980** kriged analyses of the predictions and observations of sulfur wet deposition (kg S/ha).	E-4
E.6	Areas of significant under/overpredictions of **annual 1980** amounts of sulfur wet deposition.	E-8

LIST OF FIGURES

Figure No.	Title	Page No.
E.7	Annual 1980 locations of centroids.	E-11
E.8	Annual 1980 locations and magnitudes (kg S/ha) of the maximum amount of sulfur wet deposition at the ISDME sites.	E-11
E.9	Comparison of the annual 1980 observations and predictions of the air concentrations ($\mu g/m^3$) of sulfur dioxide and sulfate at the APN sites.	E-12

LIST OF TABLES

Table No.	Title	Page No.
1.1	The models and participants in the ISDME.	5
2.1	Attributes of the statistical models as applied in the ISDME.	9
2.2	Attributes of the deterministic models as applied in the ISDME.	13
3.1	The effects on the 1980 AES predictions of sulfur wet deposition as a consequence of using precipitation data from the precipitation chemistry sites as opposed to the grids.	25
3.2	The UDDBC and ISDME overall data quality levels.	34
3.3	The distribution of sites according to site representativeness level.	36
3.4	The number of sites in the 1980 ISDME evaluation data set for each season and the year.	39
3.5	The identification codes and coordinates of the ISDME sites.	41
3.6	Mean air concentrations of sulfur dioxide and sulfate at the four APN sites during 1980.	45
4.1	Elements of the evaluation of the ISDME models.	49
4.2	The ISDME model performance measures.	52
4.3	The criteria values of the primary model performance measures used to cluster the models.	63
5.1	Comparison of kriged variances for 1980 annual amounts of observed and predicted sulfur wet deposition.	74
6.1	Summary of the predictions provided by each model.	75
6.2	Types of patterns of significant differences for **summer 1980.**	92
6.3	Summary of model performance measures for **summer 1980.**	93
6.4	Types of patterns of significant differences for **spring 1980.**	110
6.5	Summary of model performance measures for **spring 1980.**	111

LIST OF TABLES

Table No.	Title	Page No.
6.6	Comparison of the geographical regions where each model predicts the highest amounts of sulfur wet deposition for **summer/spring** versus **autumn 1980**.	124
6.7	Types of patterns of significant differences for **autumn 1980**.	128
6.8	Summary of model performance measures for **autumn 1980**.	129
6.9	Types of patterns of significant differences for **winter 1980**.	145
6.10	Summary of model performance measures for **winter 1980**.	146
7.1	The seasonal, overall and annual model performance scores.	156
A.1	Summary of previous evaluations of regional deposition models.	A- 2
B.1	The annual 1980 emission rates of sulfur dioxide (ktonnes/yr) and precision estimates for Canada.	B-11
B.2	The annual 1980 emission rates of sulfur dioxide (ktonnes/yr) for the United States.	B-12
B.3	The distribution of sulfur dioxide emissions from United States and Canadian point sources by physical stack height.	B-13
C.1	Criteria for unrepresentative monitoring sites.	C- 5
C.2	Data completeness rating criteria for seasonal sulfate values.	C- 8
C.3	Data completeness rating criteria for annual sulfate values.	C- 3
C.4	Differences in the seasonal and annual 1980 amounts of sulfur wet deposition (expressed as percent) at colocated or closely located samplers.	C-12
C.5	Differences in the annual amounts of sulfur wet deposition (expressed as percent) at colocated or closely located samplers for 1981-1982.	C-13
D.1	Comparison of the observations and predictions of **annual 1980** sulfur wet deposition at the ISDME evaluation sites.	D- 2
D.2	Comparison of the observations and predictions of **winter 1980** sulfur wet deposition at the ISDME evaluation sites.	D- 4

LIST OF TABLES

Table No.	Title	Page No.
D.3	Comparison of the observations and predictions of **spring 1980** sulfur wet deposition (kg S/ha) at the ISDME evaluation sites.	D- 6
D.4	Comparison of the observations and predictions of **summer 1980** sulfur wet deposition (kg S/ha) at the ISDME evaluation sites.	D- 8
D.5	Comparison of the observations and predictions of **autumn 1980** sulfur wet deposition (kg S/ha) at the ISDME evaluation sites.	D-10
D.6	Standardized residuals of **annual 1980** sulfur wet deposition at the ISDME evaluation sites.	D-12
D.7	Standardized residuals of **winter 1980** sulfur wet deposition at the ISDME evaluation sites.	D-14
D.8	Standardized residuals of **spring 1980** sulfur wet deposition at the ISDME evaluation sites.	D-16
D.9	Standardized residuals of **summer 1980** sulfur wet deposition at the ISDME evaluation sites.	D-18
D.10	Standardized residuals of **autumn 1980** sulfur wet deposition at the ISDME evaluation sites.	D-20
D.11	Statistical comparison of the observations and predictions of **annual 1980** sulfur wet deposition.	D-22
D.12	Statistical comparison of the observations and predictions of **winter 1980** sulfur wet deposition.	D-23
D.13	Statistical comparison of the observations and predictions of **spring 1980** sulfur wet deposition.	D-24
D.14	Statistical comparison of the observations and predictions of **summer 1980** sulfur wet deposition.	D-25
D.15	Statistical comparison of the observations and predictions of **autumn 1980** sulfur wet deposition.	D-26
E.1	Summary of model performance measures for the **annual 1980** period.	E-13

ACKNOWLEDGMENTS

The authors wish to acknowledge the efforts of all of those who contributed to the International Sulfur Deposition Model Evaluation project, especially the modelers and their colleagues who have provided model results. Without their assistance, this report would not have been possible. The following people have also contributed to this study:

Ms. Carmen Benkovitz, Ms. Veronica Evans and Mr. Marty Leach, Brookhaven National Laboratory

Dr. Perry Samson and Mr. Jerry Keeler, University of Michigan

Dr. Maris Lusis, Ontario Ministry of the Environment

Dr. Anthony Olsen, Battelle, Pacific Northwest Laboratory

Mr. Dale Coventry, Mr. Brian Eder, and Dr. Peter Finkelstein, U.S. Environmental Protection Agency

Ms. Debbie Palka and Ms. Debbie Bieler, Program Resources Incorporated

Dr. Bernard Fisher, Central Electricity Research Laboratories, United Kingdom

Mr. Dan McNaughton, TRC Environmental Consultants

EXECUTIVE SUMMARY

Computer models have been developed to estimate mean air concentrations of sulfur dioxide and sulfate and sulfur dry and wet deposition amounts at receptor points or across grid cells approximately 100 km on a side. Many of these models have been applied to estimate source-receptor relationships, particularly for sulfur wet deposition, and total sulfur deposition across geopolitical entities. In the past, these models could not be rigorously evaluated for their predictions of sulfur wet deposition, because of the paucity of sampling data. The increased number of sites in the North American precipitation chemistry networks after 1979 affords the opportunity to more rigorously evaluate these models for sulfur wet deposition.

Recognizing the need to evaluate existing North American models, the Office of Research and Development of the U. S. Environmental Protection Agency and Atmospheric Environment Service of Environment Canada jointly conducted the International Sulfur Deposition Model Evaluation (ISDME) Project. Using unique methods, this project evaluated eleven linear-chemistry atmospheric models of sulfur deposition for each season and the year of 1980. These models were either statistical, Lagrangian or hybrids. The selection of 1980 as the evaluation year was based, in part, on the availability of the most recent emissions inventory.

The approach of the study was to i) compile and distribute standardized model input data sets to the modelers, ii) apply the eleven models using these data sets and iii) evaluate the models in a "blind" test mode, meaning the computer code of these models was not modified specifically to improve the results for 1980. The evaluation data consisted of sulfur wet deposition amounts calculated from screened precipitation chemistry data from the five major North American networks and daily air concentrations of sulfur dioxide and sulfate at four Canadian sites. Dry deposition data were not available during this study.

The ISDME screening and calculation procedures were very similar to those recommended by the Unified Deposition Data Base Committee established by the Canadian Research and Monitoring Coordinating Committee and the U.S. National Acid Precipitation Assessment Program. The number of sites with data passing the ISDME criteria ranged from 32 for the annual to 46 for the spring periods.

Unlike the evaluations of the past, the ISDME focused on the ability of the models to *replicate the spatial patterns* of *seasonal*, as well as annual, amounts of sulfur wet deposition *within the uncertainties of the data*. Seasonal and annual evaluations, rather than only annual evaluations, were conducted to provide a more stringent test of the models. Patterns were generated by interpolating the site observations and predictions to half-degree grid cells via a technique known as kriging. This technique minimizes the interpolation errors and quantifies the uncertainties arising from both the interpolation and the errors of measurement.

The emphasis on pattern replication is based on the fact that point measurements are not necessarily representative of the spatial resolution of the models. That is, large spatial variability of meteorological parameters and air concentrations can cause "local" effects on the measurements at a site. As a result, these measurements could be representative of only a point and not the region. Therefore, it is necessary to compare predictions and observations on commensurate spatial scales. Moreover, pattern comparisons provide at least a limited evaluation of the models in data sparse regions.

The second unique aspect of the evaluation approach was the consideration of the uncertainties of the observed amounts of sulfur wet deposition. Rather than comparing predictions to the best estimates of the observations, this evaluation determined whether and by how much the predictions were outside the uncertainty limits of the interpolated observations. Based on these uncertainty limits, the statistical significance of the differences was determined. These uncertainties arise from sampling and analytical errors, local or nonregional effects, network protocol differences, and the interpolations of both the predictions and observations. The uncertainties of the annual pattern of sulfur wet deposition ranged from approximately ± 50% in the region of highest sulfur wet deposition to ± 100% across data-sparse regions of eastern North America.

The models were evaluated using six model performance measures, which quantified the differences between the predicted and observed patterns of sulfur wet deposition. The measure receiving the most weight was the percentage of the half-degree cells where the predictions significantly differed from the measurements. The remaining five measures were assigned equal weights. Two of these measures quantified differences in pattern position, while two additional measures quantified the differences in the magnitudes and locations of the maxima. The final measure quantified differences in the seasonal distribution of the annual amounts.

The comparison of predicted and observed patterns of seasonal sulfur wet deposition revealed that the predicted patterns tended to resemble concentric ellipses and to show less detail than the observed patterns. The predicted seasonal and annual patterns also consistently exhibited maximum amounts within the high emissions region of Ohio, western Pennsylvania and West Virginia. This behavior was not evident in the patterns of the observations, which indicated that the location of the maximum was not anchored, but migrated from northern West Virginia to southern Ontario.

Depending on the model, distances separating the predicted and observed locations of the maxima for all seasons were less than 170 to 350 km. Most models were within the uncertainty limits of the magnitude of the seasonal and annual maxima.

The determination of the areas of statistically significant differences between the predictions and observations of sulfur wet deposition revealed that across 80% of the evaluation region:

- The seasonal predictions of five of the eleven models were within the uncertainty limits of the observed seasonal patterns
- The seasonal predictions of all but three models were within the same uncertainty limits for at least three seasons
- The annual predictions of all but three models were within the uncertainty limits of the observed annual pattern.

A contributing factor to this high percentage was the degree of uncertainty of the observed pattern of sulfur wet deposition, which was as high as ± 50% across the area of greatest site density. Therefore, the models in this study were afforded a rather large statistical tolerance for differences.

Regarding the seasonal contributions to the 1980 amount of sulfur wet deposition, nearly 40% and 30% of the annual observed amounts occurred during summer (July, August and September) and spring (April, May and June), respectively. Approximately 15% occurred during both winter (January, February and March) and autumn (October, November and December). Most of the models are within 10% of the summer and spring contributions and within 25% of the autumn and winter contributions. Three models did not mimic the observed seasonal variability as well as the other models.

Two features of model performance were common to the statistical models but not to the deterministic models. First, the patterns of the sulfur wet deposition, as determined by the statistical models, tended to be oriented more toward the east-west axis, as opposed to the southwest-northeast axis of the observed patterns. On the other hand, the orientation of the patterns produced by the deterministic models tended to be closer to that of the observed patterns. This distinction could be due to the fact that the statistical models base their transport simulation on seasonal mean winds, which in North America tend to be westerly, rather than on wind measurements available twice a day. The second feature common to statistical models was the tendency to predict sulfate air concentrations higher than those of the deterministic models at both the ISDME sites and the four Canadian monitoring sites.

Finally, the seasonal, annual and overall performances of the models were summarized by scoring and clustering procedures based on subjectively determined criteria values for each of the six performance measures. A model received points if the values of the performance measures did not exceed the criteria values. Overall performance scores were based on weighted seasonal performance scores. The weights, the sum of which equalled one, were determined by the mean relative contributions of the observed seasonal amounts of sulfur wet deposition. Consequently, the model performance for the summer, the season of greatest sulfur wet deposition, received a weight of approximately 0.4. The distribution of overall performance scores illustrated three distinct clusters or groups of models. One model having one of the highest overall performance scores was deemed to have performed relatively well, but used an empirical "correction" (or background) term accounting for 40 to 60% of the predicted sulfur wet deposition.

In conclusion, this study demonstrated the use of a unique approach to model evaluation, which accounts for the uncertainties of the evaluation data. This approach is appropriate for future evaluations of the recently developed Eulerian models of wet deposition. By applying this unique approach, this study assessed the performance of linear-chemistry atmospheric models of sulfur deposition to identify the strengths and weaknesses of each model.

It is recommended that the reasons explaining the behavior of these models be investigated as a means of identifying and implementing scientifically defensible improvements to the models. It is further recommended that the improved models be evaluated for subsequent years when emissions data are available and more precipitation chemistry data exist. Additional sites would reduce the interpolation uncertainties and thus provide for a more rigorous model evaluation.

1. INTRODUCTION

1.1 Background

Within the last ten years, regional long-term sulfur deposition models have been developed to estimate monthly, seasonal, and annual air concentrations of sulfur dioxide and sulfate and wet/dry or total sulfur deposition amounts across Europe and North America with spatial resolutions ranging from 80 to 150 km. The results of some of these models, which do not usually concur, have been used to estimate source region-receptor region relationships (hereafter referred to as source-receptor relationships) on regional and international scales. These relationships have been used to investigate emission control strategies to reduce sulfur deposition amounts in sensitive regions of Europe and North America and to improve visibility by reducing sulfate air concentrations in the United States.

Since at this time it is impossible to validate these source-receptor relationships, the models should at least be evaluated for consistency with long-term regional patterns using available air concentration and wet deposition measurements of sulfur compounds and a rigorous statistical protocol. One could reason that if a model cannot replicate long-term regional patterns, its defined source-receptor relationships would be less credible. This is not to say, however, that a model which is capable of replicating the observed patterns is also capable of accurately defining source-receptor relationships. One would first need to substantiate that the model is producing the right answer for the right reasons. However, because of severe data limitations, this is difficult to do.

Not only should these models be evaluated via a rigorous statistical protocol, but the performance of each model should be intercompared to illustrate the similarities and differences and to attempt to link performance with model approach. In past evaluation studies, intercomparisons were infeasible for several reasons. First, with few exceptions, common meteorological, emissions, and evaluation data and evaluation protocol were not used in the independent studies (see Appendix A). Secondly, in North America, model evaluations were limited to the eastern portion of the continent and to periods ranging from a few days to a few months. The short evaluation periods and the restricted geographic region reflected the limitations of available evaluation data bases and, in some cases, the difficulty in generating long-term model input data bases and model results.

The first North American model evaluations focused on air concentrations, since:

1) An extensive data base for daily sulfur dioxide and sulfate air concentrations was compiled during the 1977 - 1978 SUlfur Regional Experiment (SURE)

intensive sampling periods in the northeastern United States (Perhac, 1978).

2) No dry deposition measurements were available for the pertinent temporal and spatial scales.

3) Sulfur wet deposition measurements were sparse.

The few operational sites of the Multi-State Atmospheric Power Production Pollution Study (MAP3S) and the CANadian Sampling of Air Pollution (CANSAP) networks provided sulfur wet deposition data across parts of the northeastern United States and eastern Canada during the same period. However, the small number of sites precluded a statistically meaningful evaluation for sulfur wet deposition amounts (Schiermeier and Misra, 1983).

By 1980, however, the number of precipitation chemistry sites in North America significantly increased as a result of the expansion of existing networks and the establishment of additional networks. The new networks were the United States National Atmospheric Deposition Program (NADP) network, the Electric Power Research Institute (EPRI) network, the Canadian Air and Precipitation Monitoring Network (APN) and the Acid Precipitation In Ontario Study (APIOS). The significant increase in the number of wet deposition monitoring sites now made it feasible to evaluate regional models for sulfur wet deposition amounts.

Recognizing the need to evaluate existing regional sulfur deposition models on a common basis and to determine whether these models satisfactorily replicate sulfur wet deposition patterns, the Office of Research and Development of the U.S. Environmental Protection Agency and the Atmospheric Environment Service of Environment Canada conducted the International Sulfur Deposition Model Evaluation (ISDME). This study evaluated eleven regional models using seasonal sulfur wet deposition amounts calculated from screened 1980 precipitation chemistry data.

1.2 ISDME Goals

The ISDME goals have been tailored to the needs of both the modeler and the model user (e.g., policy analyst). These goals also reflect the experience gained from past model evaluations and the perceived limitations of the model input and available evaluation data.

For the modeler, the results of a rigorous model evaluation can identify strengths and weaknesses of a particular version of a model. From this and other information, the modeler can re-evaluate and refine the model, since model development is a continuous, evolutionary process relying on iterations of both model modifications and evaluations.

For the model user, specifically, the policy analyst who focuses on the predicted source-receptor relationships, a thorough model evaluation can establish the reliability of a model. This brings us closer to the point of assessing the confidence of its source-receptor relationships.

It is recognized that a thorough model evaluation should not only investigate whether a model produces the *right results*, but it also should address the question, did the model produce the right results for the *right reasons*? This question needs to be answered to establish a level of confidence one would have that a model which produced the right results for one scenario would also produce the right results for another scenario.

The unavailability of air concentration measurements across North America, however, precluded a thorough model evaluation. Therefore, it is not always possible to substantiate whether those models producing the right results did so for the right reasons. Furthermore, the lack of a credible means of identifying source-receptor relationships precluded an evaluation of the modeled relationships. On the other hand, with the available wet deposition data, we can assess the performance of the models in estimating sulfur wet deposition.

The specific goals of the ISDME are as follows:

1) Evaluate operational sulfur deposition models for each season for an entire year.

2) Identify which models best replicate the observed spatial and temporal patterns of sulfur wet deposition within data uncertainty limits.

3) Delineate for each model, areas and periods of good and poor agreement with the observed sulfur wet deposition amounts.

4) Intercompare predictions of air concentrations and dry deposition amounts.

5) Investigate the effects on model performance of the uncertainties of processed precipitation data grids and the observed sulfur wet deposition amounts.

6) Attempt to explain the model performances.

Originally, the ISDME goals included an intercomparison of model source-receptor matrices. However, due to the time constraints of the modelers, this task was deleted.

To achieve these goals, special investigations were first conducted to explore methods and techniques to:

1) Screen the precipitation chemistry measurements for completeness and the sites for regional representativeness;

2) Calculate sulfur wet deposition amounts;

3) Spatially analyze predictions and observations of sulfur wet deposition with minimal interpolation errors;

4) Quantify the uncertainties of the measurements and the interpolations; and

5) Rigorously evaluate and intercompare the performance of the models, given the limitations of the evaluation data.

The results of these investigations, discussed in subsequent sections and appendices of this report, are useful in future model evaluation studies. In addition, a special investigation conducted by Clark et al. (1986) assessed the significance of the bias observed in the precipitation grids on the predictions of sulfur wet deposition and the performance of one of the ISDME models.

1.3 ISDME Participants

The criteria considered in selecting models for this evaluation were as follows:

1) The model must be capable of simulating seasonal sulfur dioxide and sulfate air concentrations and sulfur wet and dry deposition amounts across eastern North America with a resolution of at least 150 km and

2) By June of 1985, the modeler must be able to submit model results for all seasons using the 1980 ISDME standardized model input data set.

Letters inviting modelers to participate were mailed to those modelers participating in the U.S./Canadian Memorandum of Intent on Transboundary Air Pollution and authors of published scientific articles concerning regional sulfur deposition models. The nine participating modelers and the eleven ISDME models are listed in Table 1.1.

TABLE 1.1 THE MODELS AND PARTICIPANTS IN THE ISDME.

Model Name	Participant and Affiliation
Atmospheric Environment Service Long-Range Transport Model (AES)	Eva Voldner and Marvin Olson Environment Canada
Advanced Statistical Trajectory Regional Air Pollution (ASTRAP) Model	Jack Shannon Argonne National Laboratory
Ontario Ministry of the Environment (MOE) Long-Range Transport Model	Barbara Ley Ministry of the Environment
Regional Impacts on Visibility and Acid Deposition (RIVAD) Model	Doug Latimer Systems Applications Inc.
Regional Lagrangian Model of Air Pollution (RELMAP)	Terry Clark U.S. Environmental Protection Agency
Regional Transport Model (RTM-II)	Doug Latimer Systems Applications Inc.
Statistical Estimates of Regional Transport and Acidic Deposition (SERTAD) Model	Ron Portelli Concord Scientific Corp.
Statistical Model (STATMOD)*	Akula Venkatram Environmental Research & Technology Inc.
Tennessee Valley Authority (TVA) Model-T	William Norris Tennessee Valley Authority
University of Michigan Atmospheric Contributions to Interregional Deposition (UMACID) Model	Perry Samson University of Michigan

---

* A second version of this model, SIMPMOD, was also applied.

1.4 Structure of the Report

The remaining portion of this report is structured as follows:

- Chapter 2 highlights the attributes of each of the ISDME models.
- Chapter 3 discusses the ISDME standardized model input data set and the sulfur wet deposition and air concentration data bases.
- Chapter 4 describes the ISDME model evaluation criteria and the measures to quantify model performance.
- Chapter 5 describes the interpolation technique applied to spatially analyze sulfur wet deposition amounts.
- Chapter 6 presents and discusses the values of the performance measures for each model, intercompares model performances and intercompares model results for those parameters for which appropriate model evaluation data are unavailable.
- Chapter 7 compares the model performances for each season and discusses the overall performance of each model.
- Chapter 8 presents the conclusions of the study.
- Finally, Chapter 9 offers recommendations for future research.

Numerous appendices discuss the results of special investigations or provide details not included in one of the above chapters. The text in the main body of the report sometimes references figures and tables found in one of the appendices. These are distinguished from the figures and tables appearing in the main body by the use of the letter of the Appendix where they are located. For example, Figure B.1 is the first figure of Appendix B.

2. MODEL DESCRIPTIONS

The regional models evaluated in this study are designed to predict average air concentrations and total wet/dry deposition amounts of sulfur dioxide and sulfate for periods of one month to a year. Since their primary objective is to assess control strategies on groups of sources or emission regions, the spatial scale of the models is such that point and area sources within 80 to 150 km grid cells are aggregated into single virtual sources.

Emissions of sulfur dioxide from these virtual sources and the sulfate resulting from chemical transformation are subjected to vertical mixing and horizontal diffusion at rates prescribed by the model as the pollutants are transported across North America. During transport, percentages of the pollutant mass are dry deposited and scavenged by precipitation according to rates prescribed by the model and the precipitation data used by the model. A portion of the total mass is transported beyond the continental shores, off the model grid, and is no longer considered.

Some models consider a background level of air concentrations attributed to biogenic and distant anthropogenic sources. As a result, these models enhance their predictions of dry and wet deposition.

Sulfur deposition models can be classified as being either purely statistical, Lagrangian, Eulerian (ordered here on the basis of increasing levels of sophistication) or hybrids of the three. The fundamental differences in these three model types, elaborated by Pasquill and Smith (1983), relate to the approaches of simulating the physical and chemical processes.

In simple terms, the four ISDME statistical models (Table 2.1) generally simulate transport and diffusion via one or more long-term average, single-level wind roses, which represent speed and directional frequencies for 22.5 degree sectors. Wet deposition of pollutant mass is typically dependent on the average frequency and duration of wet and dry periods rather than the actual occurrences.

On the other hand, the seven ISDME Lagrangian or deterministic models (Table 2.2) simulate the transport of pollutant puffs via objectively analyzed winds at one or more levels. Physical and chemical processes are simulated for each puff. Horizontal diffusion is simulated in a variety of approaches based on puff travel time and/or diffusion coefficients and/or ensemble puff trajectory fluctuations. Typically, the wet deposition of pollutant mass is dependent on the rate and time of occurrence of precipitation and pollutant air concentrations. Ensemble concentrations and deposition amounts are determined by adding the contributions of each puff.

Finally, Eulerian models simulate transport and diffusion, as well as other processes, across a geographically fixed area, rather than computing ensemble averages and totals. None of the ISDME models were purely Eulerian.

Basic attributes of the ISDME models are only highlighted here. More detailed information can be obtained from the modelers whose names appear in Table 1.1. These models can be compared to European and other North American models in Pasquill and Smith (1983) and the U.S. Environmental Protection Agency (1987).

TABLE 2.1 ATTRIBUTES OF THE STATISTICAL MODELS AS APPLIED IN THE ISDME.

		SERTAD	TVA	SIMPMOD/STATMOD
a.	Model Type	Statistical	Statistical	Statistical
b.	Receptor Configuration	Receptor points	Receptor points	Receptor points
c.	Model Domain	North America and adjacent oceans	E. North America and adjacent ocean	North America
d.	Model Output	Seasonal-annual SO_2 and $SO_4^=$ concentrations and wet depositions and SO_2 dry depositions	Seasonal-annual SO_2 and $SO_4^=$ concentrations and dry/wet depositions	Seasonal-annual SO_2 and $SO_4^=$ concentrations and dry/wet depositions
e.	Model Input Data			
	Emissions	Seasonal SO_2 point and area sources on 127-km grid[1]	Seasonal SO_2 point and area sources on 127-km grid[1]	Seasonal SO_2 point and area sources on 127-km grid[1]
	Wind, Temperature and Pressure	6-h 1000-mb winds on 127-km grid[1]	6-h 1000 and 850-mb winds on 127-km grid[1]	6-h 850-mb winds on 127-km grid[1]
	Precipitation	12-h amounts on 127-km grid[1]	1-h amounts on 127-km grid[1]	3-h amounts on 127-km grid[2]

[1] ISDME standardized data set
[2] SIMPMOD uses no precipitation data

TABLE 2.1 CONTINUED

	SERTAD	TVA	SIMPMOD/STATMOD
f. Treatment of Emissions	Uniformly distributed across 127-km grid cells	Uniformly distributed across 127-km grid cells	Uniformly distributed across 127-km grid cells
g. Calculation of Transport Wind	16-sector wind roses computed across E. North America from surface winds on 127-km grid	1 16-sector wind rose computed by averaging mixed layer winds at 64 127-km grid cells[2]	Seasonal and annual 850-mb, 16-direction wind roses computed and averaged for 22 climatological regions
h. Treatment of Precipitation	Total amount[1] and duration of dry and wet periods computed at each 127-km grid cell for each period	Mean length of dry and wet periods and mean rates during wet periods computed for 64 127-km grid cells[2]	Annual mean rates computed for each 127-km grid cell[3]
i. Mixed Layer (m)	Two seasonal mixing heights for entire domain: 374-498 for stable and 677-1199 for unstable	Spatially constant: 702 (winter) 1212 (spring) 1276 (summer) 831 (autumn) 1005 (annual)	Seasonal and annual climatological heights on 127-km grid[1]
j. Horizontal Dispersion	360° spreading weighted by directional frequency and diluted by mean wind speed in direction of receptor	360° spreading weighted by directional frequency and diluted by mean wind speed in direction of receptor	360° spreading weighted by directional frequency and diluted by mean wind speed in direction of receptor

[1] ISDME standardized data set
[2] The selection of the 64 cells was based on the location of the evaluation sites (Chapter 3)
[3] Not used by SIMPMOD

TABLE 2.1 CONTINUED

	SERTAD	TVA	SIMPMOD/STATMOD
k. Vertical Dispersion	Gradient theory with constant eddy diffusivity for SO_2; instantaneous mixing for $SO_4^=$	Uniform throughout mixed layer	Uniform throughout mixed layer
l. Transformation Rate (% h^{-})	Varies seasonally: 0.67 (winter) 1.0C (spring/ autumn/ annual) 1.33 (summer); S as Sulfate in emission cell(%): 20.0 (winter) 30.0 (spring/ autumn/ annual) 40.0 (summer)	Varies seasonally: 0.28 (winter) 0.80 (spring) 0.85 (summer) 0.34 (autumn) 0.57 (annual)	Varies seasonally: 0.7 (winter) 1.3 (spring) 1.5 (summer) 0.7 (autumn) 1.0 (annual)
m. Dry Deposition Velocity (cm s^{-1})			
SO_2	Varies seasonally: 0.33 (winter) 0.50 (spring/ autumn/ annual) 0.60 (summer)	0.5	Varies seasonally: 0.3 (winter) 0.6 (spring) 0.7 (summer) 0.4 (autumn) 0.5 (annual)
$SO_4^=$	Not considered	0.05	Varies seasonally: 0.04 (winter)
		0.05 (spring)	
			0.06 (summer) 0.05 (autumn) 0.05 (annual)

TABLE 2.1 CONTINUED

		SERTAD	TVA	SIMPMOD/STATMOD
n.	Wet Removal Rate (% h^{-1})			
	SO_2	24 (winter) 36 (otherwise)	Varies seasonally: 66.0 (winter) 72.4 (spring) 76.0 (summer) 65.5 (autumn) 70.9 (annual)	Varies seasonally: 7.2 (winter) 10.8 (spring) 10.8 (summer) 9.0 (autumn) 9.9 (annual)
	$SO_4^=$	24 (winter) 36 (otherwise)	Varies seasonally: 68.0 (winter) 72.4 (spring) 76.0 (summer) 65.5 (autumn) 70.9 (annual)	Varies seasonally: 23.4 (winter) 36.0 (spring) 36.0 (summer) 29.9 (autumn) 31.3 (annual)
o.	Background Deposition (kg $SO_4^=$ ha^{-1})	Function of spatial precipitation pattern and period: c P_T, where P_T is precipitation amount for season/year (mm) and c (mg $SO_4^=$ l^{-1}) varies seasonally: 0.000 (winter) 1.688 (spring) 2.250 (summer) 2.625 (autumn) 2.124 (annual)	None	None

TABLE 2.2 ATTRIBUTES OF THE DETERMINISTIC MODELS AS APPLIED IN THE ISDME.

		AES	ASTRAP	MOE	RELMAP
a.	Model Type	Lagrangian box	Lagrangian puff; Statistical trajectory	Lagrangian puff	Lagrangian puff
b.	Receptor Configuration	Receptor points or 52 x 58 grid cells; 127 km at 60° N.	51 x 45 grid cells; 127 km at 60° N.	37 x 38 grid cells; 127 km at 60° N.	45 x 30, 1° x 1° grid cells; 80 x 110 km at 40° N.
c.	Model Domain	North America and adjacent oceans	North America and adjacent oceans	E. North America and adjacent ocean	E. North America and adjacent ocean
d.	Model Ouput	Daily-annual SO_2 and $SO_4^=$ concentrations and dry/wet depositions	Monthly-annual SO_2 and $SO_4^=$ concentrations and dry depositions and bulk S wet depositions	Daily-annual SO_2 and $SO_4^=$ concentrations and dry/wet S depositions	Daily/annual SO_2 and $SO_4^=$ concentrations and dry/wet depositions
e.	Model Input Data				
	Emissions	Seasonal SO_2 point and area sources on 127-km grid[1]	Seasonal SO_2 point and area sources on 3-dimensional 127-km grid with 100-m vertical spacing[2] ; stack parameters	Seasonal SO_2 point and area sources on 127-km grid[1]	Seasonal SO_2 point and area sources on 1° x 1° 3-layer grid[1]
	Wind, Temperature, and Pressure	6-h winds at 1000 and 850 mb on 381-km grid[1]	6-h winds at 1000 and 850 mb on 127-km grid[1]	3-h MSL pressures at synoptic sites	12-h winds at 1000 and 850 mb on 1° x 1° grid[1]
	Precipitation	24-h amounts on 127-km grid[1]	1-h amounts on 127-km grid[1]	12-h amounts on 127-km grid[1] and 3-h amounts at synoptic sites	12-h amounts on 1° x 1° grid[1]

[1] ISDME standardized data set

[2] ISDME standardized data set with minor variation

TABLE 2.2 CONTINUED

	AES	ASTRAP	MOE	RELMAP
f. Treatment of Emissions	Added to moving box every 3 h; mixed instantaneously in mixed layer	Released in puffs; diffused from source layer at 6-h intervals	Released in puffs at 3-h intervals; mixed instantaneously in mixed layer	Released in puffs at 12-h intervals; initial mass released into 1-3 layers; instantaneously mixed during day; no nocturnal mixing
g. Calculation of Transport Wind	Linear interpolation of 1000- and 850-mb winds	$1/d^2$ weighting of 1000 and 850-mb winds to produce 2-D fields for 6 layers up to 3 km on 381-km grid	3-h MSL pressures analyzed via orthogonal polynomials to calculate 3-h geostrophic winds	1000-mb winds in lowest layer; 850-mb winds in top layer; 1:4 weighting of surface:850-mb winds in middle layer
h. Treatment of Precipitation	No analysis necessary	1-h amounts summed for 6-h periods on 127-km grid	3-h amounts determined by apportionment of 12-h amounts based on supplementary data for each 127-km grid cell	12-h amounts assumed to occur in one 3-h period for each 1° x 1° grid cell
i. Mixed Layer (m)	Monthly climatological maxima/minima on 127-km grid[1]	Diurnally varies with eddy diffusivity profile; maxima: 1000 (winter) 1800 (spring/summer) 1400 (autumn)	Varies seasonally: 500 (winter) 800 (spring/autumn) 1000 (summer)	Monthly climatological maxima/minima on 1° x 1° grid[1]

[1] ISDME standardized data set

TABLE 2.2 CONTINUED

	AES	ASTRAP	MOE	RELMAP
j. Horizontal Dispersion	Emissions evenly distributed across grid cells; long-term dispersion by spatial/temporal distributions of trajectories	Long-term dispersion by fitting distributional functions to ensemble of trajectories from each virtual source for 1-3 months	Gaussian distributions about trajectory end points: $\sigma = (\sigma_0{}^2 + k_d t)^{0.5}$, $\sigma_0 = 32$ km and $k_d = 5 \times 10^5 m^2 s^{-1}$	Constant at 339 km^2 h^{-1}
k. Vertical Dispersion	Uniform throughout mixed layer	Varies diurnally and seasonally according to eddy diffusivity profiles in 1-D numerical integration; flux across top of mixed layer	Uniform throughout the mixed layer	Uniform throughout the mixed layer during the day; No nocturnal mixing
l. Transformation Rate (% h^{-1})	Varies monthly: 0.6 (January) 1.4 (July) 1.0 (annual mean)	Varies diurnally and seasonally: 0.16-1.12 (winter) 0.32-3.20 (spring) 0.40-4.00 (summer) 0.24-2.00 (autumn)	Varies seasonally: 0.5 (winter) 1.0 (spring/autumn) 2.0 (summer)	Varies diurnally, seasonally and latitudinally; at solar room at 45° N.:0.36(winter) 1.19 (spring/autumn) 1.94 (summer)
m. Dry Deposition Velocity (cm s^{-1})				
1. SO_2	Varies monthly: 0.4 (January) 0.8 (July) 0.6 (annual mean)	Varies diurnally/ seasonally: 0.10-0.70 (winter) 0.10-0.80 (spring) 0.10-0.90 (summer) 0.10-0.65 (autumn)	Varies seasonally: 0.5 (winter/spring/ autumn) 0.8 (summer)	Varies seasonally with stability index and land-use type: 0.05-0.45 (winter) 0.05-1.15 (spring/ summer) 0.05-0.85 (autumn)

TABLE 2.2 CONTINUED

	AES	ASTRAP	MOE	RELMAP
m. Dry Deposition Velocity (cm s^{-1})				
$SO_4^=$	Varies monthly: 0.08 (January) 0.10 (July) 0.09 (annual mean)	Varies diurnally and seasonally: 0.05-0.25 (winter) 0.05-0.40 (spring) 0.05-0.45 (summer) 0.05-0.36 (autumn)	Varies seasonally: 0.05 (winter/spring/ (autumn) 0.08 (summer)	Varies seasonally with stability index and land-use type: 0.05-0.25 (winter) 0.10-0.40 (spring/ (autumn) 0.10-0.50 (summer)
n. Wet Removal Rate (% h^{-1})*				
SO_2	$3 P_{24}/H$	Bulk S removal rate (% per 6 h): minimum $\left\{\begin{array}{l}3.16 P_6^{0.5}\\ 50\end{array}\right.$ but north of 36° N in winter: minimum $\left\{\begin{array}{l}1.58 P_6^{0.5}\\ 50\end{array}\right.$	$10.8 P_1$	Varies seasonally: $a P_3^b$ where $0.01 < a < 0.14$ $0.12 < b < 0.70$
$SO_4^=$	$100 P_{24}^{0.6}/H$		$108 P_1$	Varies seasonally: $c P_3^d$ where $0.02 < c < 0.39$ $0.06 < d < 0.70$
o. Background Deposition	None	None	2.0 kg S ha^{-1} yr^{-1} everywhere	None

* P_n refers to precipitation rates (mm per n hours); H refers to mixing heights (km).

TABLE 2.2 CONTINUED

		RIVAD	RTM-II	UMACID
a.	Model Type	Lagrangian plume segment	Hybrid Eulerian/ Lagrangian puff	Lagrangian puff
b.	Receptor Configuration	45 x 36 grid cells; 80 km	45 x 36 grid cells; 80 km	receptor points
c.	Model Domain	North America and adjacent oceans	North America and adjacent oceans	E. North America and adjacent ocean
d.	Model Output	Daily-annual SO_2 and $SO_4^=$ concentrations and dry/wet depositions	Daily-annual SO_2 and $SO_4^=$ concentrations and dry/wet depositions	Daily-annual SO_2 and $SO_4^=$ concentrations and dry/wet depositions
e.	Model Input Data			
	Emissions	Seasonal SO_2 sources on 80-km grid [1]	Seasonal SO_2 sources on 80-km grid [1]	Seasonal SO_2 sources on latitude/longitude grid [1] where Δx = 1° and Δy =Δx at 42° N
	Wind, Temperature and Pressure	6-h winds at 1000 and 850 mb on 80-km grid [1]	6-h winds at 1000 and 850 mb on 80-km grid [1]	12-h wind profiles
	Precipitation	12-h amounts on 127-km grid [1]	12-h amounts on 127-km grid [1]	1-h amounts on 127-km grid

[1] ISDME standardized data set

TABLE 2.2 CONTINUED

		RIVAD	RTM-II	UMACID
f.	Treatment of Emissions	Released as plume segments from major point sources at effective stack height at 3-h intervals; release height of area source emissions dependent on stability	Released as Gaussian puffs from major point sources at effective stack heights; all other emissions instantaneously mixed in mixed layer	1 and 5% of S is $SO_4^=$ in winter and summer, respectively
g.	Calculation of Transport Wind	Wind at plume segment centroid height interpolated between data levels and between grid nodes	Mixed layer and upper winds determined from weighting 1000- and 850-mb winds	3-h winds computed using data within 300 nm of trajectory segment end point using $1/d^2$ weighting
h.	Treatment of Precipitation	12-h amounts assumed to occur in one 3-h period for each 80-km grid cell	12-h amounts assumed to occur in one 3-h period for each 80-km grid cell	1-h amounts summed and gridded onto latitude/longitude grids
i.	Mixed Layer (m)	Varies diurnally; monthly climatological maxima/minima on 80-km grid[1]	Varies diurnally; monthly climatological maxima/minima on 80-km grid[1]	Varies monthly

---

[1] ISDME standardized data set

TABLE 2.2 CONTINUED

		RIVAD	RTM-II	UMACID
j.	Horizontal Dispersion	Gaussian: $\sigma_y \alpha t^{1.2}$, $t < 10h$ $\sigma_y \alpha t^{0.5}$, $t > 10h$	K_h dependence on wind field deformation; imposed maximum/minimum; σ_y for Gaussian submodel	Gaussian: $\sigma_x = \sigma_y = 1.8t$
k.	Vertical Dispersion	Dependent on stability: $\sigma \alpha t^k$ where k varies from 0.3 for F to 2.1 for A	Gaussian in puff submodel; constant flux layer within mixed layer; flux between layers	Uniform throughout mixed layer
l.	Transformation Rate ($\% h^{-1}$)	Varies diurnally and seasonally: 0.1-0.2 (winter) 0.1-1.0 (spring/autumn) 0.1-1.5 (summer)	Varies diurnally, seasonally and latitudinally: 0.5-1.5 (winter) 0.5-1.5 (spring) 0.5-2.0 (summer) 0.5-1.6 (autumn)	Varies diurnally and seasonally: 0.1-1.4 (winter) 0.2-2.8 (summer)
m.	Dry Deposition Velocity ($cm\ s^{-1}$)			
	SO_2	Varies with surface type; 0.14-1.50 during day; 0.00 at night	Varies with meteorological parameters and surface type; 0.85-1.00 during day; 0.0C at night	Varies diurnally and seasonally: 0.10-0.55 (winter) 0.10-0.82 (summer)

TABLE 2.2 CONTINUED

		RIVAD	RTM-II	UMACID
m.	Dry Deposition Velocity (cm s^{-1})			
	$SO_4^=$	Varies with surface type; 0.2-0.5	Varies with meteorological parameters and surface type; 0.35-1.00 during day	Varies diurnally and seasonally: 0.05-0.28 (winter) 0.03-0.43 (summer)
n.	Wet Removal Rate (% h^{-1})*			
	SO_2	$\Lambda_a P_3/H$ ‡	$\Lambda_a P_3/H$ ‡	$5\ P_1/H$
	$SO_4^=$	$\Lambda_b P_3^{0.625}/H$ ‡	$\Lambda_b P_3^{0.625}/H$ ‡	$29\ P_1^{0.622}$
o.	Background Deposition	None	None	None

---

* P_n refers to precipitation rates (mm per n hours); H refers to mixing heights (km).

‡ Λ_a and Λ_b dependent on the ratio of the concent^ation in aqueous phase to that in air.

3. ISDME DATA SETS

3.1 Standardized Model Input Data

Since models can be sensitive to the input data (i.e., the emissions and meteorological data), sets of standardized data grids were compiled and distributed to the modelers to minimize the differences in model results due solely to differences in the data sources or data processing methods. Whenever more than one source of the data existed, the most appropriate source was selected. As described later, data from these selected sources were screened for potential errors.

As with any measurement or estimate, uncertainties are inevitable. Assessing the uncertainties of the model predictions arising from the uncertainties in the input data was not within the scope of this study. However, as discussed in Chapter 5, uncertainties in the sulfur wet deposition data were considered when evaluating the models.

This chapter discusses the processing methodologies of the input data provided to the participating modelers and the screening and calculation procedures to compile the evaluation data base.

3.1.1 General Considerations

In the process of selecting a period for regional deposition model evaluation one must consider the availability of quality controlled wet deposition data and model input data (i.e., wind, precipitation and emissions data). Meteorological data, which have been collected via networks with rather constant resolutions over many years, do not constrain the evaluation to certain years. On the other hand, appropriate emissions and wet deposition data are not available for many years. Wet deposition data were not available at a sufficient number of sites until 1980, while the most recent year for which detailed North American sulfur dioxide emissions data are available is 1980. Consequently, 1980 was selected as the year for model evaluation. Essential data for this year were acquired and processed to create a standardized set of model input data.

Processing of model input data involved the creation of standardized data grids that were compatible to the ISDME models (Clark et al., 1985). The universal application of one set of data grids is complicated by the fact that models do not always use the same grid configuration or spatial resolution. Since the grid configuration is an integral part of the model, model input data grids were created for each of the four grid configurations used by the ISDME models as illustrated in Figure B.1. The spatial resolution of the grids, which varies with latitude and longitude, averaged from 80 to

127 km.

Most regional sulfur deposition models require wind velocity grids for at least one layer of the boundary layer, high resolution precipitation grids and sulfur dioxide emission grids. These are the major components of the ISDME standardized data set. Grids of mean monthly maximum and minimum mixing heights are also included.

3.1.2 Transport Winds

Upper-air winds are measured only twice per day (early morning and early evening) at sites spaced approximately 350 km apart, except across northern Canada where distances are greater. Thus, errors are introduced when interpolating the data to finer resolutions (e.g., 6-h grids with an 80-km resolution for several levels of the boundary layer).

On the other hand, wind analyses from dynamic meteorological models offer a better alternative for several reasons. First, since the dynamic meteorological models include three-dimensional parameterizations of pertinent physical processes and a time history of the data, the analyses between measurement sites and for intermediate periods is likely to be more accurate than analyses derived solely from horizontal interpolations. Secondly, these dynamic models screen the wind, temperature and height data before using the values in the analyses by examining measured wind profiles and using time series to identify measurements that are inconsistent with physical laws. Finally, satellite data are included in the data assimilation process to compensate for data paucity in some areas.

On the basis of these assets, it was decided that dynamic model winds would be used for the ISDME. Since they were readily available from archived tapes, 6-h 381-km wind grids produced for the 1000- and 850-mb levels by the Canadian Meteorological Centre (CMC) primitive equation, spectral model were selected for ISDME model applications. The wind analysis method is described by Rutherford (1977). These 381-km wind grids were reduced to the finer resolution ISDME configurations via a triangularization technique (Lawson, 1972).

Quality control measures consisted of comparing the finer resolution ISDME wind velocity contour plots to the 381-km wind velocity plots for each 6-h interval of one day in each season. Maximum and minimum speeds for each interval were scanned for each grid configuration to identify any questionable extremes. High wind speeds were verified using National Climatic Data Center upper-air charts. No significant inconsistencies were detected.

3.1.3 Precipitation Amounts

There are several archived precipitation data bases of varying temporal and spatial resolutions for North America. Since the timing, duration and amount of precipitation could be important factors in simulating wet deposition amounts, one would prefer to use the finest spatial and temporal resolution precipitation data base to derive model input data grids. The data with the highest resolution is the hourly North American precipitation data base, which consists of measurements from nearly 2000 sites. Unfortunately, unresolved data inconsistencies and spatial voids prohibited their use.

The major problems with this data set were discovered after grids of hourly precipitation amounts were generated. With no explicit indication, data were either frequently missing and/or the reported amounts at many sites were much less than actual amounts. Monthly grids generated by summing hourly amounts for all sites in each grid cell detected numerous cells with unrealistically low amounts, differing by an order of magnitude from the amounts at nearby 24-h network sites and at adjacent cells. The cause of this problem was traced to the data themselves and not the data processing procedure. Another problem was the sparse coverage in Canada, especially in northern Canada during the snow season when many of the sites are inoperable.

As an alternative to this data base, the 24-h 1980 objectively analyzed, CMC precipitation grids, with a resolution of 127 km at 60 degrees North latitude, were selected. Despite the coarser temporal and spatial resolutions, this data base was more attractive for several reasons. First, the data were already in a grid form. Secondly, unrealistic monthly totals were not detected for any cell. Finally, amounts for data sparse or data void areas, estimated from ship data or data from nearby sites, were already assimilated into the data base. The precipitation analysis method is described by Davies (1967).

One drawback of these data grids, however, is the consequential spatial smoothing of the analysis method. As a result, the spatial variability of the precipitation amounts decreases. Theoretically, this would tend to smooth the model results as well.

Since many models require 6-h precipitation amounts, the 24-h amounts in each of the CMC grid cells were resolved to shorter periods. Since the 24-h periods in this data base overlapped the preceding and succeeding periods by 12 hours, 12-h amounts could be inferred. As the average precipitation event in the northeastern United States ranged from 2.5 to 3.9 hours, depending on the season (Thorp and Scott, 1982), the 12-h amounts were assumed to have occurred over a 3-h period beginning at the midpoint of the 12-h period (i.e., 06-09 or 18-21 GMT).

Although they are the only existing data, McNaughton (1984) concluded that the precipitation data from conventional hourly and daily networks do not adequately represent those precipitation amounts required by the models. The relatively sparse network forces modelers to apply spatial and temporal averaging techniques, which contribute to the uncertainty of the predictions of sulfur wet deposition. He further concluded that by reducing the precipitation grid dimensions from 80 to 40 km, a factor of two uncertainty in deposition predictions could arise as a consequence of the increased significance of the gaps in the coverage of precipitation monitoring sites.

Since the dimensions of the ISDME precipitation grids ranged from 80 km to approximately 130 km, the significance of the gaps is not expected to be as great. Nonetheless, uncertainties do exist because of the averaging techniques employed by the model preprocessors.

As a means of identifying any bias that might have been introduced in spatially averaging precipitation data, the 1980 annual precipitation amounts at 72 eastern North American precipitation chemistry sites, not used in creating the CMC analyses, were compared with the annual amounts at the 127-km CMC grid cells containing each site. Figures B.2-B.6 compare the patterns of 1980 seasonal and annual precipitation amounts derived from the CMC grids and the measurements at the precipitation chemistry sites.

The patterns are somewhat similar in shape and position, but the magnitudes of the grid cell amounts appear to be less than the site amounts. Ruff et al. (1984) reported a similar observation in another study. The histograms of Figure B.7 verify the lower grid values: The annual medians for the grid and site amounts were 815 and 910 mm, respectively. For the grid amounts, the first and third quartiles were much closer to the median indicating less variability than was found in the measurements. This can be attributed in part to the fact that data averaging operations to derive precipitation amounts for a given area smooth the measurements.

Differences in the grid cell and site precipitation data sets are further illustrated by the 1980 weekly time series plots for 18 ISDME sites (Figure B.8). At these sites, much of the annual bias in the grid cell precipitation amounts can be explained by the bias for only a few weeks primarily in spring and summer. During these seasons, one would expect that the larger biases reflect the increased spatial variability of both the precipitation occurrences and rates due to the domimance of convective versus stratified precipitation events. Since predicted deposition amounts are a function of precipitation occurrences and rates, the uncertainties of the grid cell precipitation would be expected to be the greatest in spring and summer.

The effects of these uncertainties can be assessed by substituting precipitation amounts measured at the precipitation chemistry sites for the spatially averaged precipitation amounts conventionally used by the models. For each season and the year of 1980, Clark et al. (1986) have assessed these effects by comparing AES model predictions to observations of sulfur wet deposition at 18 precipitation chemistry sites in eastern North America (Figure B.8). The AES model was selected because of the availability of its weekly predictions based on the CMC precipitation amounts.

The results of this comparison, presented in Table 3.1, indicate that for every season, as well as for the year, both the mean predictions and the explained variances increased as a consequence of using precipitation data measured at the precipitation chemistry sites. The explained variances increased by at least 65% and, with the exception of autumn, mean predicted amounts increased by about 10% or less. This increase in the predictions is a rather small amount, but it nonetheless is indicative of one component of the total uncertainty. Because there are differences in the manner in which each model preprocesses data and computes wet deposition amounts, the

TABLE 3.1 THE EFFECTS ON THE 1980 AES PREDICTIONS OF SULFUR WET DEPOSITION AS A CONSEQUENCE OF USING PRECIPITATION DATA FROM THE PRECIPITATION CHEMISTRY SITES AS OPPOSED TO THE GRIDS.

PERIOD	INCREASE IN MEAN PREDICTIONS (%)	VARIANCE EXPLAINED SITE	GRID	INCREASE (%)
WINTER	4	0.12	0.20	65
SPRING	10	0.18	0.42	129
SUMMER	10	0.31	0.61	94
AUTUMN	37	<0.01	0.06	2204
ANNUAL	11	0.10	0.25	160

uncertainties attributed to the averaging techniques will vary.

3.1.4 Climatological Mixing Heights

Some models consider spatial, diurnal and seasonal variations in mixing heights. Since there is no consensus on the objective method to determine mixing heights from rawinsonde data, only climatological North American mixing height grids were included in the standardized data set.

These maximum and minimum mixing height grids were based on January, April, July and October mixing heights for early evening (00 GMT) and early morning (12 GMT) periods as determined by Holzworth (1967 and 1972) and Portelli (1977) from several years of rawinsonde data. Grids generated for these 4 months were linearly interpolated to create grids for the remaining months. A constant 200-m mixing height was assumed by Holzworth and Portelli over large bodies of water.

3.1.5 Anthropogenic Sulfur Dioxide Emissions Data

Since the ISDME models assume linear chemistry, the magnitude of the uncertainty in the emissions input data will be directly reflected in the predictions of sulfur wet deposition. That is, if the emissions data were overestimated everywhere by 10%, the predictions of sulfur wet deposition everywhere would be 10% higher than they would be if the emissions data were not overestimated. This example illustrates the importance of an accurate emissions data base.

3.1.5.1 Canadian Data

The Canadian sulfur dioxide emissions inventory compiled by the Canadian Environmental Protection Service (EPS) for the MOI Emissions, Costs, and Engineering Assessment Work Group 3-B (MOI, 1982b) was selected for the ISDME, since it represented the most appropriate Canadian inventory. The data in this inventory pertained to the period from 1978 through 1980. However, approximately 85% of the point source emissions (accounting for 80% of the Canadian total) pertained to 1980.

Emissions from all other anthropogenic sources were summed to area source totals. Area source emissions, comprising 6% of Canada's total emissions, represented 1978. However, according to Vena (private communications, 1984), these area source emission rates are not expected to vary significantly from 1980 levels.

Point source data, available on a stack basis, included stack location, height, diameter, effluent temperature, flow rate, seasonal and annual emission rates, etc. Seasonal area source emissions, on the other hand, were available on the 127-km subgrid of the 381-km CMC grid. For the emissions data, seasons were defined as follows:

Winter:	December, January and February
Spring:	March, April and May
Summer:	June, July and August
Autumn:	September, October and November.

The total sulfur dioxide emission rates and the EPS precision estimates for each Canadian province are listed in Table B.1. Sulfate emissions, estimated to represent 2.7% of the Canadian sulfur emissions total, were assimilated into the sulfur dioxide inventory.

The precision estimates were determined using a three-step approach. First, a check for bias errors or omissions was undertaken and corrections were performed when required. Secondly, systematic errors were then determined on a provincial basis for each major point source and for each major point source category, based on an engineering analysis of the parameters that influence the computed emissions. Thirdly, a weighted sensitivity analysis was performed to calculate the precision of the nationwide and provincial total emissions. The overall precision of the Canadian sulfur dioxide emissions total is estimated to be approximately ±6% (MOI, 1982b).

3.1.5.2 United States Data

Subsequent to the MOI project, the Man-Made Sources Task Group of the National Acid Precipitation Assessment Program revised the emissions inventory used in developing the MOI data base. The data were revised in several stages as updated source information and advances in emission estimation methodologies became available. At the time a data base was needed for the ISDME, Version 2.0 of the NAPAP inventory was available.

The Version 2.0 annual national emissions rate was 4.7% higher than that for the MOI inventory. The revised total was a consequence of the improvements of the MOI inventory, which included

1) Updated 1980 sulfur dioxide emissions data from twelve states,

2) Updated EPA sulfur dioxide emission factors,

3) Updated 1980 power plant emissions data from the U.S. Department of Energy and

4) Disaggregated annual area source emission rates to seasonal rates.

More recently, Version 5.0 of the NAPAP emissions inventory was completed. A comparison of the Version 2.0 and Version 5.0 inventories indicated that

1) The annual sulfur dioxide emission rates across the 48 contiguous states was 0.4 % greater in the Version 2.0 inventory and

2) The annual sulfur dioxide emission rates across the northeastern industrialized states differed by less than 6%.

Like the Canadian inventory, NAPAP point source data were available on a stack basis for each season, as defined for the Canadian inventory. By definition, a United States point source was a stationary source emitting at least 100 tons per year of any of the five primary criteria pollutants. Approximately 94% of the nation's 1980 sulfur dioxide emissions were released by point sources.

Not all point source data pertained to one year. However, over 95% of the nation's annual sulfur dioxide emissions from point sources pertained to 1980. Much of the remaining emissions data pertained to years preceding 1980.

Seasonal area source emission rates were available for 64 source categories for each county (or air pollution control district, as was the case for Massachusetts, Connecticut and Maine). Total emission rates are listed by state in Table B.2. These sulfur dioxide emission rates account for primary sulfate emissions, which contribute an estimated 1.6% of the national annual sulfur emissions (MOI, 1982b). The precision of the Version 2.0 inventory was not estimated, since it was an interim inventory. The precision estimates of the state emission rates in the MOI emissions inventory would not apply here, since many data records were updated.

3.1.5.3 Aggregation Methods

For each ISDME grid configuration and for each season, the Canadian and United States point source emissions were aggregated to three-dimensional grid cells according to the location of the stacks and their estimated effective stack heights. Due to the numerous instances of missing or incorrect stack coordinates and stack heights in the United States inventory, approximations were used when necessary to determine locations and effective stack heights.

Discrepancies in the stack coordinates or missing coordinates were encountered for approximately 20% of the nearly 200,000 United States point sources. These discrepancies, which accounted for approximately 10% of the total United States sulfur dioxide emissions from point sources, were identified by comparing the inventory source coordinates with the coordinates along the border of the smallest rectangle encompassing the county containing the stack. If the inventory stack location was found to be outside this rectangle, the coordinates of the county centroid replaced the inventory source coordinates.

The problem of missing stack heights was approached by using the average stack heights computed for each of the Standard Industrial Classification (SIC) codes in the NAPAP inventory. Approximately 25% of the sulfur dioxide point sources had incomplete stack parameters accounting for approximately 15% of the total United States sulfur dioxide emissions.

United States and Canadian point source emissions were aggregated to grid cells containing the sources. Since pollutant transport and dispersion are dependent on release height, the point source emissions were further aggregated to three layers, 0-200, 201-500 and greater than 500 m, after doubling the physical stack heights (Figure 3.1). These layers were selected so that approximately the same emissions would be emitted in each of the lowest two layers and that the emissions from the very tall stacks would be separate. The frequency distribution of the effective stack heights and the total emissions for 50-m layers are listed in Table B.3.

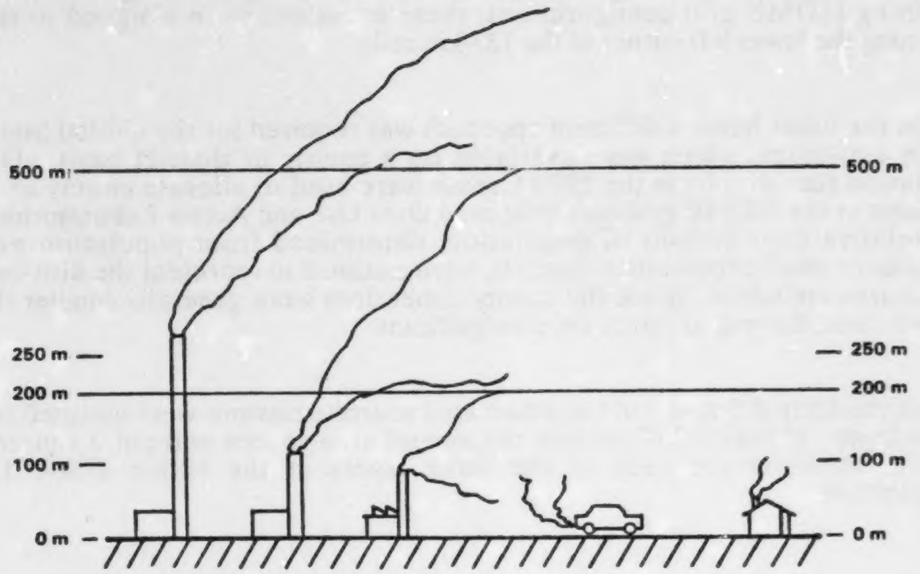


Figure 3.1 The three layers into which sulfur dioxide emissions from point and area sources are injected.

Release heights were estimated for dominant source categories and representative stack parameters. Climatological data and Briggs' plume rise equations were used (Briggs, 1969, 1971 and 1972). The United States emissions, predominantly from power plants, were subdivided into three categories based on physical stack heights of 190, 220 and 340 m. Canadian emissions were subdivided into two categories of primary metal plants, pulp and paper plants, four categories of oil refineries and three categories of power plants based on physical stack heights of 91, 150 and 198 m.

An average effective stack height for each category was computed for each season using stability frequency distributions and associated mean wind speeds on a monthly basis. These analyses, derived from surface data and the STAR program (Martin and Tikvart, 1968), were available for the North American 127-km grid (Sirois and Voldner, 1987). Rather than examining each grid cell separately, the stability frequency distributions over three broad geographical regions were considered. The regions were coastal, central industrialized and southern.

The range of average stack heights over the year was found to be two to three times the physical stack heights. As a conservative estimate, an effective stack height of twice the physical stack height was chosen for all seasons.

Area source emissions were aggregated in a different manner. Canadian area source emissions were already aggregated to the 127-km grid configuration. For the remaining ISDME grid configurations, these emissions were assigned to the cells containing the lower left corner of the 127-km cell.

On the other hand, a different approach was required for the United States area source emissions, which were available on a county or district basis. Detailed population statistics from the 1970 Census were used to allocate county or district emissions to the ISDME grid cells (National Data Use and Access Laboratories, 1971). The relative distributions of population, determined from population-weighted centroids of small enumeration districts, were assumed to represent the distribution of area source emissions. Since the county dimensions were generally smaller than the grid cell sizes, the spatial errors were insignificant.

All the United States and Canadian area source emissions were assigned to the 0-200-m layer. Figure 3.2 illustrates the annual area source and point source sulfur dioxide emissions for each of the three layers of the 80-km UMACID grid configuration.

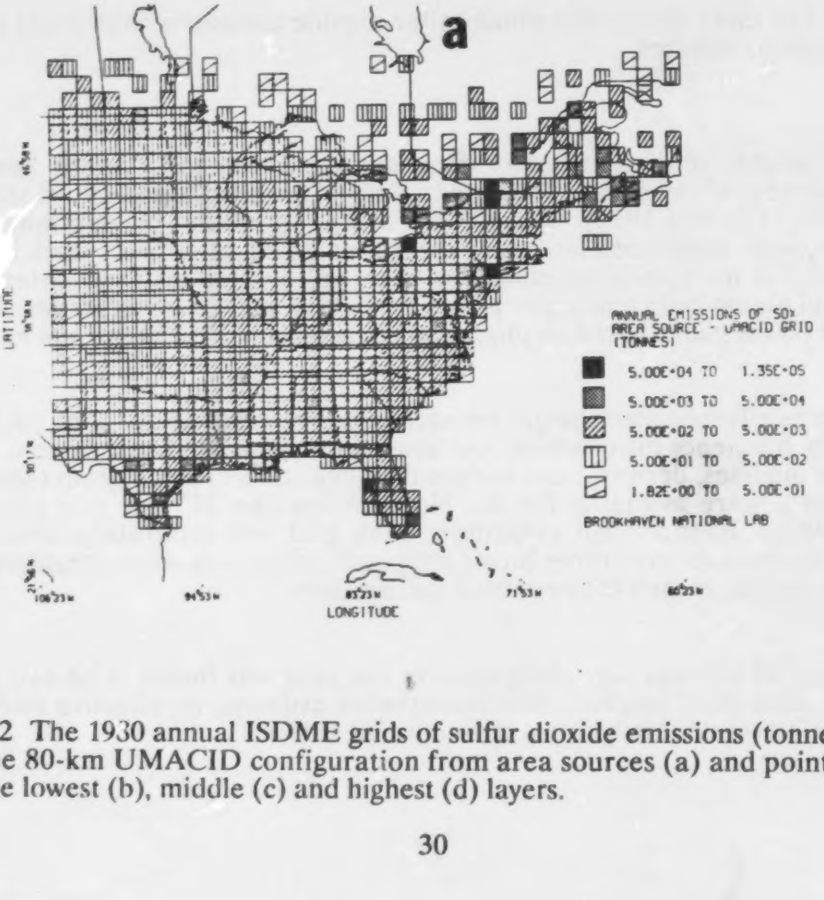


Figure 3.2 The 1930 annual ISDME grids of sulfur dioxide emissions (tonnes/year) for the 80-km UMACID configuration from area sources (a) and point sources in the lowest (b), middle (c) and highest (d) layers.

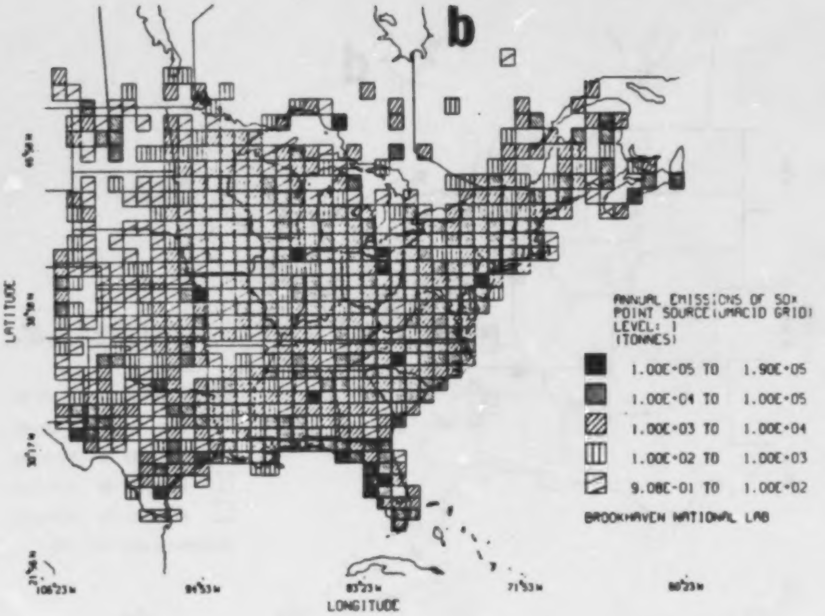


Figure 3.2 Continued.

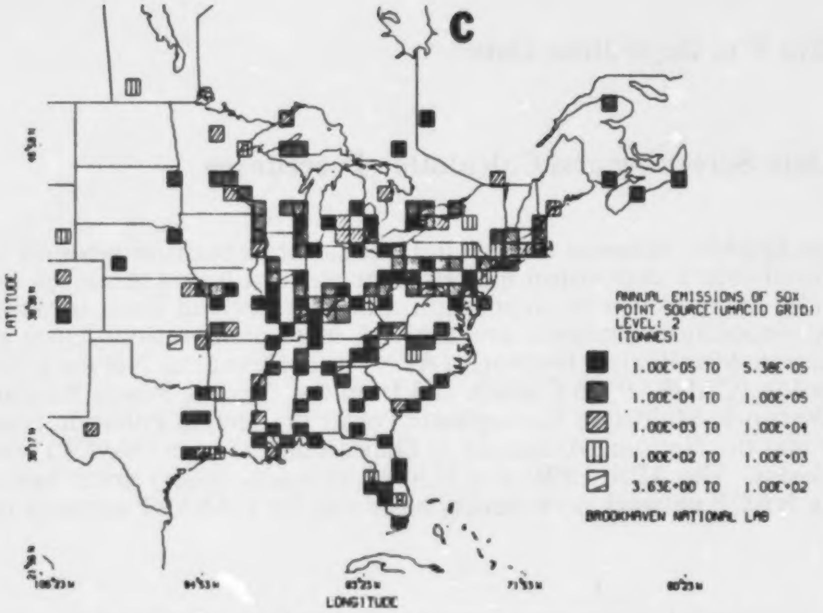


Figure 3.2 Continued.

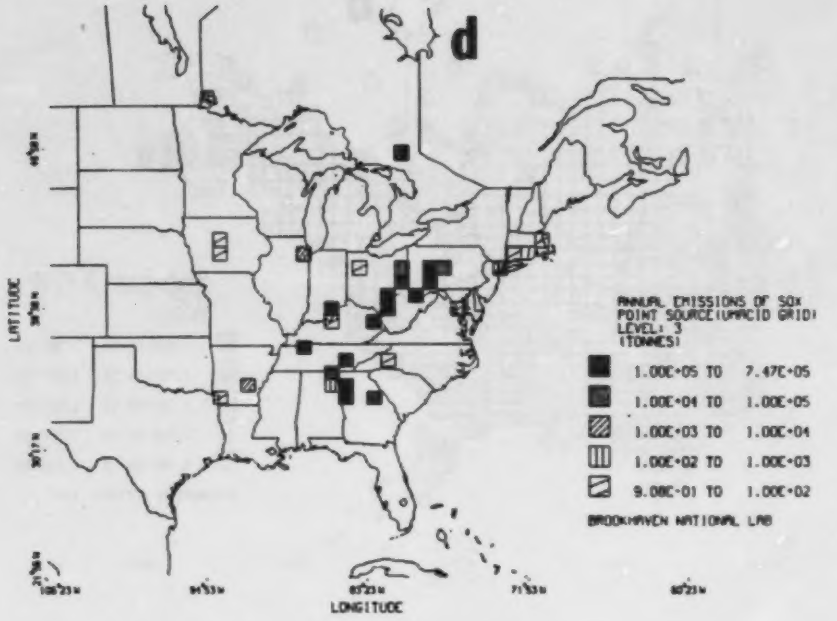


Figure 3.2 Continued.

3.2 Sulfur Wet Deposition Data

3.2.1 Data Screening and Calculation Procedures

In the ISDME, seasonal and annual wet sulfur deposition amounts predicted by the regional sulfur deposition models over eastern North America for 1980, are statistically compared with deposition amounts derived from observations. The observed deposition amounts are derived from observations from the Air and Precipitation Monitoring Network (APN) and Canadian Network for Sampling Precipitation (CANSAP) in Canada and from the Electric Power Research Institute (EPRI) Network, Multistate Atmospheric Power Production Pollution Study (MAP3S) Network and the National Atmospheric Deposition Program (NADP) Network in the United States. The APN, EPRI and MAP3S networks collect precipitation on a daily basis, the NADP network on a weekly basis and the CANSAP network on a monthly basis.

For a meaningful comparison between the predictions of the regional models and network data, it is important that the wet deposition amounts calculated from the data of the different networks are compatible, that the quality of or error in the observed amounts are known and that only regionally representative sites are considered.

Since the methods for quality control and quality assurance of the precipitation chemistry data and the procedure for computing wet deposition amounts are not always similar for the above mentioned networks, it was necessary to adopt a unified procedure for screening and correcting the data and computing deposition amounts. Also, since the wet deposition amounts estimated at the different monitoring sites may not be of the same quality, a scheme was developed to classify the data into three levels of uncertainty. The level of uncertainty is based on the deviation of the location of a monitoring site from ideal siting criteria, on the percentage of data capture, sampling efficiency and on the contamination from sea spray.

The methods of data screening/correcting, calculations of wet deposition amounts and assessment of uncertainties, originally proposed for the ISDME by Voldner et al. (1984), were refined and applied by the Unified Deposition Data Base Committee (UDDBC) established under the Canadian Federal-Provincial Research and Monitoring Co-ordinating Committee's Atmospheric Science Sub-Group (RMCC/ASSG) (Unified Deposition Data Base Committee, 1985). Under the direction of UDDBC these procedures were applied to the Acid Deposition System (ADS) data base (Watson and Olsen, 1984), the official repository of North American precipitation chemistry measurements.

With the cooperation of the UDDBC, overall data quality levels were determined and sulfur wet deposition amounts were calculated for each season and the year of 1980. The three overall data quality levels were a function of the site representativeness and data completion ratings.

Based on site characteristics and nearby sources, each site was rated as a 1, 2A, 2B or 3, where 1 is considered to be regionally representative and 3 is not, rated according to its assumed regional representativeness (Unified Deposition Data Base Committee, 1985). "Assumed" is used here to emphasize the fact that little is known about the effects of "local" sources on the precipitation chemistry. The site rating criteria are discussed in more detail in Appendix C.

In addition to the site representativeness ratings, the seasonal and annual data were rated as 1, 2 or 3 on the basis of the percentage of the data captured, sampling efficiency and the contamination from sea spray. The data with a rating of 1 is most desirable while those with a rating of 3 are considered to be an inadequate representation of the period. The data rating criteria are also discussed in Appendix C.

Table 3.2 defines the UDDBC overall data quality levels. The UDDBC recommended that only the data with an overall rating of 1 or 2 be used for model evaluations. However, the stringency of the criteria combined with the start-up of many monitors during the year resulted in an insufficient number of observations for a statistically meaningful model evaluation. Only observations at five sites in the United States passed the annual criteria for levels 1 or 2. In winter the situation was slightly better with 18 observations in the United States and 4 in Canada qualifying. Clearly, these deposition data with their uneven spatial distribution were inadequate for use in the 1980 ISDME. Thus, it was necessary to relax the UDDBC criteria and accept

TABLE 3.2 UDDBC (U) AND ISDME (I) OVERALL DATA QUALITY LEVELS.

Data Completeness Level	Site Representativeness Level					
	1		2		3	
	U	I	U	I	U	I
1	1	1	2	1	3	2
2	2	2	2	2	3	2
3	3	3	3	3	3	3

observations of potentially lower quality.

The deposition calculation procedures were not changed, however. The modified UDDBC procedures of screening the precipitation chemistry data, the procedure of calculating wet deposition amounts and an assessment of the uncertainties of the wet deposition amounts are discussed in Appendix C. The differences between the ISDME and the UDDBC's screening criteria were in the collection efficiency (ISDME criteria were lower by 5-20%) and in the overall data quality data levels. The UDDBC overall data quality level could not exceed a site's representativeness level. For the ISDME however, sites with a representativeness level of 3 were assigned overall ratings of 2 if the data completeness levels were either a 1 or 2 (see Table 3.2).

Similar to the UDDBC recommendation, only data with overall ratings of 1 or 2 were used in the ISDME. Therefore, the major difference between the evaluation data base recommended by the UDDBC and that used in the ISDME was the ISDME acceptance of those sites with a low regional representativeness rating provided that their data completeness rating was a 1 or 2.

3.2.2 Results of Data Screening and Calculations

Figure 3.3 illustrates the location of the monitors in the five networks and provides their site representativeness levels. As discussed in Appendix C, level 1 represents a "regionally representative" site, 2A and 2B "potentially representative" and "potentially unrepresentative" sites, and 3 "unrepresentative" sites. Many of the southern Ontario and eastern Canadian sites are level 3. According to Table 3.3, the majority of sites are either level 2A or 2B.

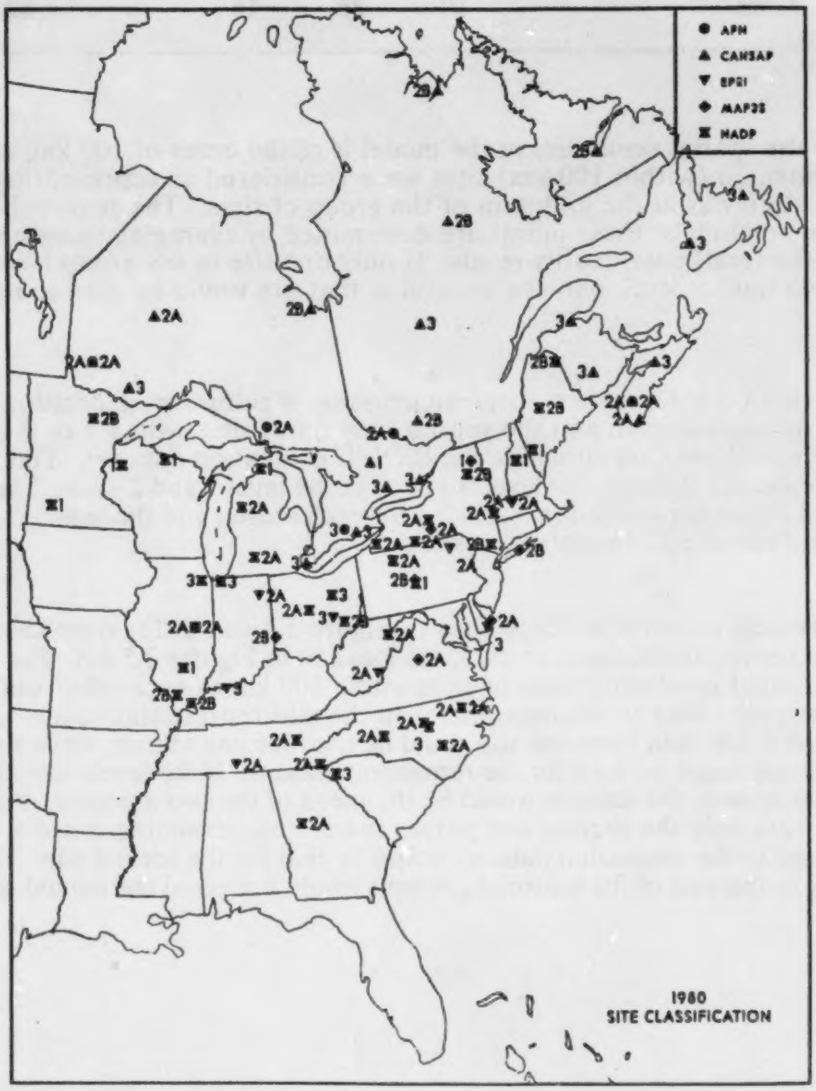


Figure 3.3 Locations and levels of regional representativeness of the precipitation chemistry sites operating in 1980 and passing at least the Level 3 criteria.

TABLE 3.3 THE DISTRIBUTION OF SITES ACCORDING TO SITE REPRESENTATIVENESS LEVEL.

	1	2A	2B	3	Total
U.S.	9	28	11	8	56
Canada	1	8	5	12	26
Total	10	36	16	20	82

Since the spatial resolution of the model is of the order of 100 km, colocated or nearly colocated (within 100 km) sites were considered as representing one point whose location was at the midpoint of the group of sites. The seasonal and annual deposition amounts at these points are determined by averaging those amounts with the highest overall data quality levels. If only one site in the group has the highest overall data quality level, only the amount at that site would be used to represent the point.

Figures 3.4-3.8 display the observed amounts of sulfur wet deposition (in units of kg S/ha) for each season and the year at only those sites with a 1 or 2 overall data quality level. These data comprise the ISDME evaluation data set. The figures also display the overall data quality level for each of the level 1 and 2 sites. The number of level 1 and 2 sites is provided in Table 3.4 for each season and the year. Depending on the period, there are 32-46 evaluation sites.

The annual amounts at some sites in Figure 3.4 can differ from the sum of the seasonal amounts as displayed at the same location in Figures 3.5-3.8. For these cases, either colocated monitoring sites or sites within 100 km of each other were operating during the year. Due to inconsistencies in the site representativeness and/or data capture levels, the data from one site could be used for one season, while the data from the second site could be used for the remaining seasons. If the levels were identical for a particular season, the amount would be the mean of the two amounts. Providing that the data from only the second site passes the annual screening criteria, the annual amount used in the evaluation data set would be that for the second site. Therefore, in this example, the sum of the seasonal amounts would not equal the annual amount.

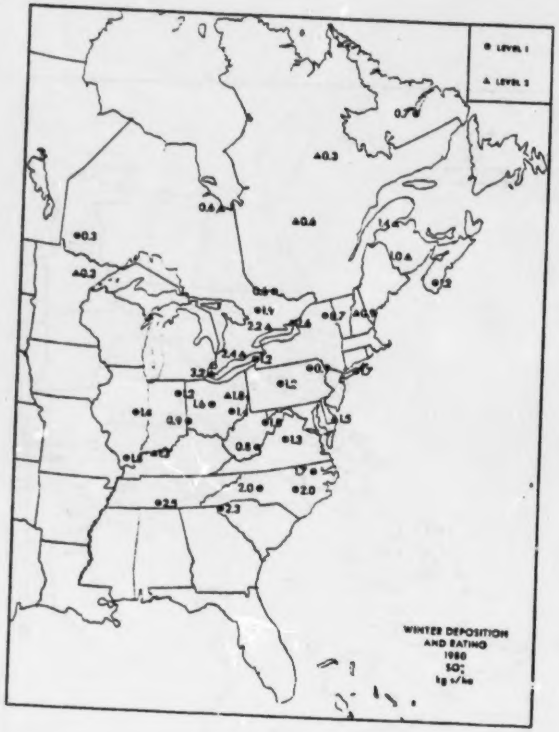


Figure 3.4 **Winter 1980** amounts of sulfur wet deposition (kg S/ha) at the ISDME evaluation sites

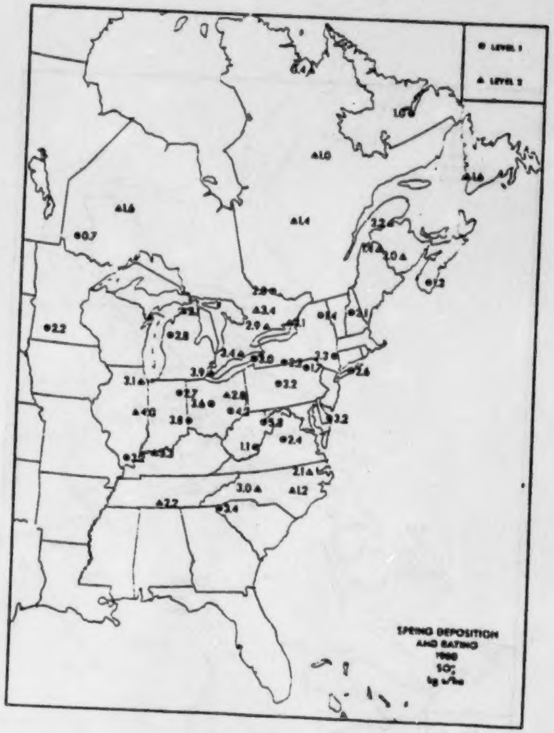


Figure 3.5 **Spring 1980** amounts of sulfur wet deposition (kg S/ha) at the ISDME evaluation sites.

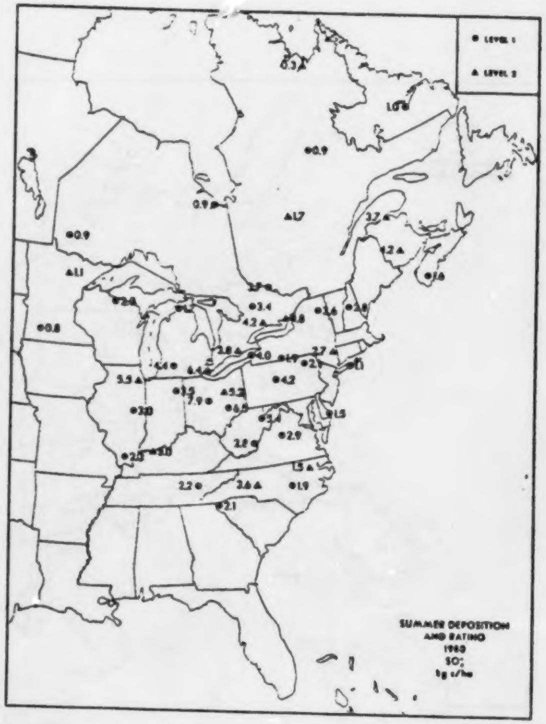


Figure 3.6 **Summer 1980** amounts of sulfur wet deposition (kg S/ha) at the ISDME evaluation sites

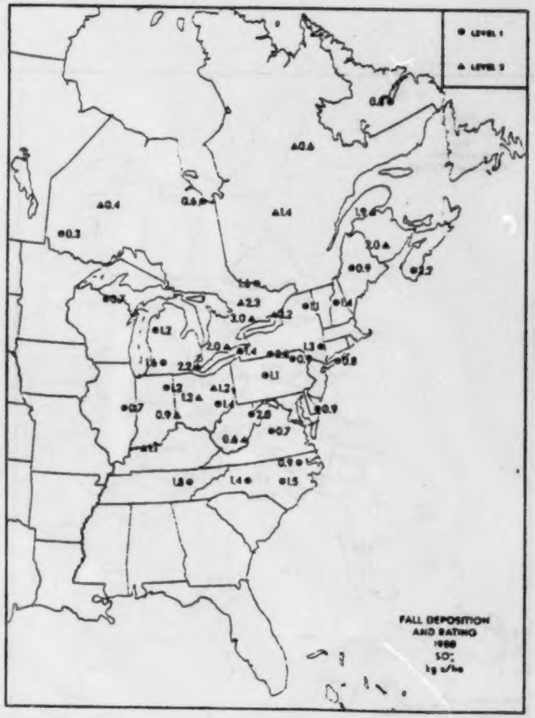


Figure 3.7 **Autumn 1980** amounts of sulfur wet deposition (kg S/ha) at the ISDME evaluation sites.

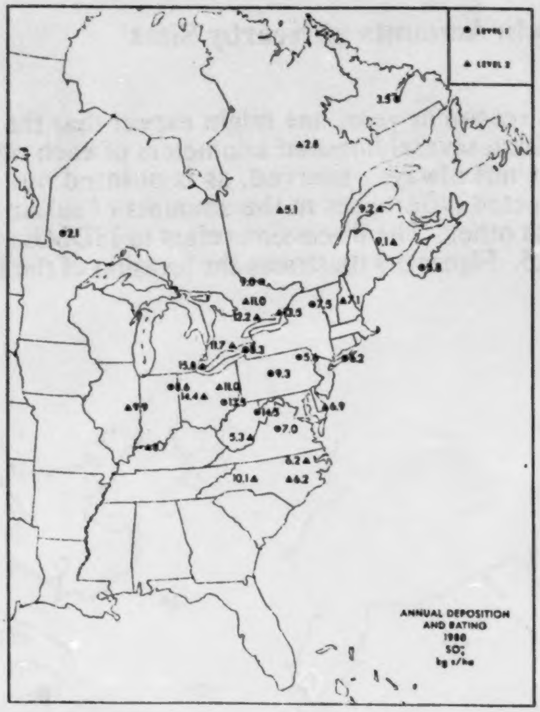


Figure 3.8 **Annual 1980 amounts of sulfur wet deposition (kg S/ha) at the ISDME evaluation sites.**

TABLE 3.4 THE NUMBER OF SITES IN THE 1980 ISDME EVALUATION DATA SET FOR EACH SEASON AND THE YEAR.

	OVERALL DATA QUALITY LEVEL		
Period	1	2	TOTAL
Winter	24	24	38
Spring	24	22	46
Summer	30	15	45
Autumn	27	15	42
Annual	13	19	32

3.2.3 Differences in Amounts at Nearby Sites

For periods of a season or year, one might expect that the amounts of sulfur wet deposition at sites within several hundred kilometers of each other would generally be comparable. This is not always observed, as is pointed out in this section, which highlights the unexpected differences in the amounts of sulfur wet deposition at sites within 200 km of each other. The discussion refers to ISDME site identification codes as defined in Table 3.5. Figure 3.9 illustrates the location of the ISDME sites.

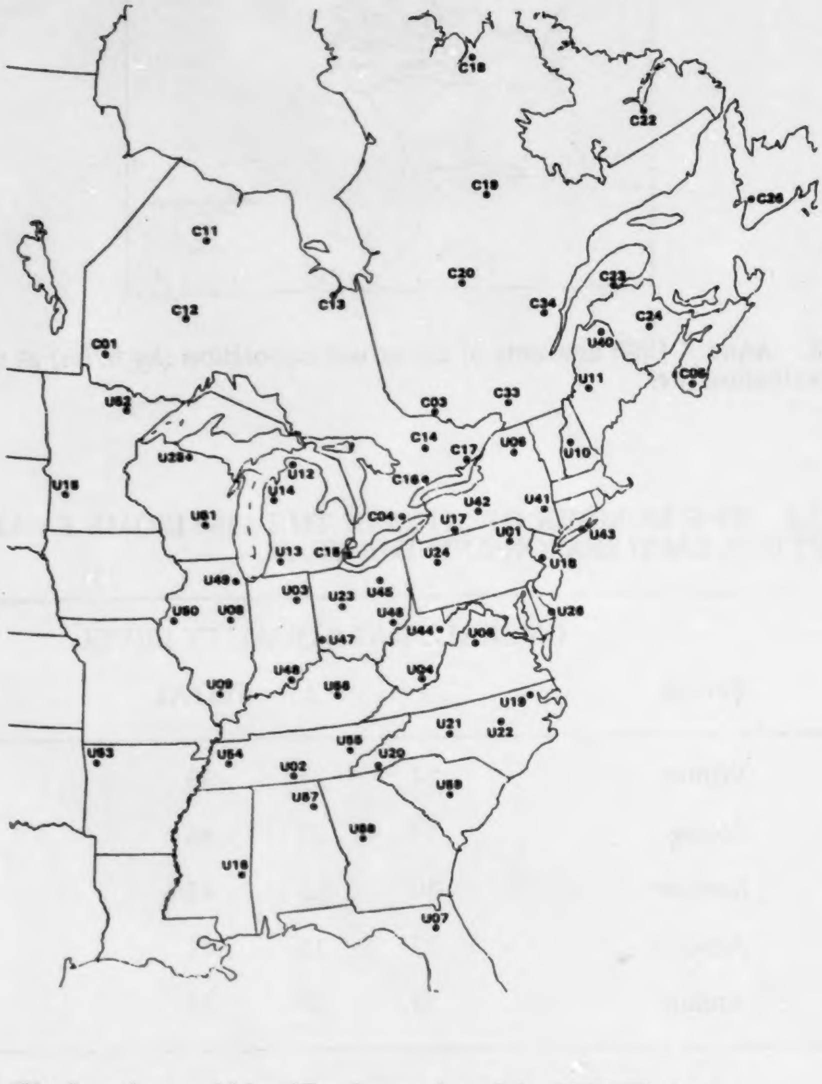


Figure 3.9 The location and identification codes of the ISDME sites.

TABLE 3.5 THE IDENTIFICATION CODES AND COORDINATES OF THE ISDME SITES.

Site Code	Site Name	Network	Latitude(°)	Longitude(°)
U01	Scranton, PA	EPRI	41.6	76.0
U02	Giles Co., TN	EPRI	35.3	86.9
U03	Fort Wayne, IN	EPRI	41.0	85.3
U04	Lewisburg, WV	EPRI	37.8	80.4
U05	*		44.2	74.1
	Whiteface, NY	MAP3S	44.4	73.9
	Huntington, NY	NADP	44.0	74.2
U06	Charlottesville, VA	MAP3S	38.0	78.5
U07	Bradford For., FL	NADP	30.0	82.2
U08	*		40.1	88.4
	Bondville, IL	NADP	40.1	88.4
	Champaign, IL	MAP3S	40.1	88.4
U09	*		37.6	89.0
	S. I. U., IL	NADP	37.7	89.3
	Dixon Springs, IL	NADP	37.4	88.7
U10	Hubbard Br., NH	NADP	43.9	71.7
U11	Greenville, ME	NADP	45.5	69.7
U12	Douglas Lake, MI	NADP	45.6	84.7
U13	Kellogg, MI	NADP	42.4	85.4
U14	Wellston, MI	NADP	44.2	85.8
U15	Lamberton, MN	NADP	44.2	95.3
U16	Meridian, MS	NADP	32.3	88.8
U17	Jasper, NY	NADP	42.1	77.5
U18	Washington Cr., NJ	NADP	40.3	74.9
U19	Lewiston, NC	NADP	36.1	77.2
U20	*		34.9	83.1
	Coweeta, NC	NADP	35.1	83.4
	Clemson, NC	NADP	34.7	82.8
U21	Piedmont St., NC	NADP	35.7	80.6
U22	*		35.5	78.6
	Clinton St., NC	NADP	35.0	78.3
	Finley A, NC	NADP	35.7	78.7
	Finley B, NC	NADP	35.7	78.7
	Raleigh, NC	EPRI	35.7	78.7
U23	Delaware, OH	NADP	40.4	83.1
U24	*		41.0	78.2
	Kane, PA	NADP	41.6	78.8
	Leading Ridge, PA	NADP	40.7	77.9
	Penn. St. U., PA	MAP3S	40.8	78.0
U25	Trout Lake, WI	NADP	46.1	89.7
U26	*		38.7	75.2
	Lewes, DE	MAP3S	38.8	75.0
	Indian River, DE	EPRI	38.6	75.3
U40	Caribou, ME	NADP	46.9	68.0

TABLE 3.5 THE IDENTIFICATION CODES AND COORDINATES OF THE ISDME SITES.

Site Code	Site Name	Network	Latitude(^{0})	Longitude(^{0})
U41	*		41.9	73.8
	Knobit, NY	NADP	42.4	73.5
	Stillwell, NY	NADP	41.4	74.0
U42	*		42.6	76.7
	Aurora, NY	NADP	42.7	76.7
	Ithica, NY	MAP3S	42.4	76.6
U43	Brookhaven, NY	MAP3S	40.9	72.9
U44	Parsons, WV	NADP	39.1	79.7
U45	Wooster, OH	NADP	40.8	81.9
U46	*		39.9	81.8
	Caldwell, OH	NADP	39.8	81.5
	Zanesville, OH	EPRI	40.0	82.0
U47	Oxford, OH	MAP3S	39.5	84.7
U48	Rockport, IN	EPRI	37.9	87.1
U49	Argonne, IL	NADP	41.7	88.0
U50	**		40.5	91.0
U51	**		43.5	89.8
U52	Marcell, MN	NADP	47.5	93.5
U53	Fayetteville, AR	NADP	36.1	94.2
U54	**		35.5	89.0
U55	Walker Br., TN	NADP	36.0	84.3
U56	**		37.5	84.0
U57	**		34.2	86.0
U58	**		35.2	84.4
C01	*		49.7	93.7
	E. L. A., ON	CANSAP	49.7	93.7
	E. L. A., ON	APN	49.7	93.7
C03	Chalk River, ON	APN	46.1	77.4
C04	*		43.0	80.6
	Simcoe, ON	CANSAP	42.9	80.3
	Long Point, ON	APN	43.0	80.8
C05	*		44.2	65.2
	Kejimkujik A, NS	CANSAP	44.4	65.2
	Kejimkujik B, NS	APN	44.4	65.2
C11	Trout Lake, ON	CANSAP	53.8	89.9
C12	Pickle Lake, ON	CANSAP	51.5	90.2
C13	Moosonee, ON	CANSAP	51.3	80.7
C14	Dorset, ON	CANSAP	45.2	78.9
C15	Harrow, ON	CANSAP	42.0	82.9
C16	Peterborough, ON	CANSAP	44.2	78.4
C17	Kingston, ON	CANSAP	44.2	76.6
C18	Fort Chimo, QU	CANSAP	58.1	68.4
C19	Nitchequon, QU	CANSAP	53.2	70.9
C20	Chibougamau, QU	CANSAP	49.8	74.4

TABLE 3.5 THE IDENTIFICATION CODES AND COORDINATES OF THE ISDME SITES.

Site Code	Site Name	Network	Latitude(°)	Longitude(°)
C23	Charlo, NB	CANSAP	48.0	66.3
C24	Acadia FES, NB	CANSAP	46.0	66.4
C26	Stephenville, NF	CANSAP	48.5	58.6
C33	**		45.5	75.0
C34	**		49.0	71.0

* Site code and coordinates pertain to either colocated monitoring sites or a point within 50 km and equidistant from two or more monitoring sites.

** No monitoring sites are located at this location; modelers were requested to provide predictions here.

As illustrated by Figure 3.4, the observed 1980 annual maximum deposition of 15.8 kg S/ha occurs at Harrow, Ontario (C15) at the western end of Lake Erie. Although this value passed the screening criteria, it did seem to be higher than expected. In 1980 the winter and spring deposition amounts at this site were considerably higher than those at the nearest sites. In the following year, the spring and summer amounts at this monthly network site were greater than those amounts at three event network sites, which were within 100 km of the former. (The latter network began operations in late 1980, hence data were not available for the ISDME). The Harrow site was classified as a level 3 site (i.e., one that is regionally unrepresentative) due to its proximity to an agricultural research station. Its operation was terminated at the end of 1982. Hence, the value should be viewed with some caution.

The seasonal and annual maps of sulfur wet deposition indicate that the amounts at some sites, even though they passed the screening criteria, consistently appeared to be lower than those amounts at the nearest sites. For most seasons, as well as for the year, the Scranton (U01) and the Lewisburg (U04) sites showed lower deposition amounts than those at sites within a radius of 200 km. Also, for the three seasons when data passed the screening criteria, the amounts at Jasper (U17) were also lower than those at the nearest sites. These apparently lower amounts cannot be explained. However, the lower amounts at the first two sites, which are EPRI network sites, could be caused in part by gaps in the monitoring periods, which are not reported in the event network data base. Also, since climatological precipitation patterns in mountainous regions show high spatial variability, these three sites, which are located in the Appalachian Mountain chain, could be located in climatologically low precipitation areas. The Lewisburg (U04) site was discontinued in 1981.

Summer and annual amounts at Kejimkujik (C05) are considerably lower than at the two New Brunswick sites (C23 and C24). These differences could reflect differences in meteorological patterns and network siting and monitoring protocol.

Although internetwork data inconsistencies have been observed here and reported in studies mentioned earlier, data adjustments were not attempted in the ISDME. Nonetheless, one should be aware that they do exist.

3.3 Air Concentration Data

For 1980, regionally representative air concentrations of both sulfur dioxide and sulfate were measured simultaneously at only four sites, CO1, CO3, CO4 and CO5. Obviously, these few measurements are not sufficient for a statistical evaluation. However, since the climatological conditions at these APN sites and their relative location to the high emission regions in North America are significantly different, a simple comparison between model predicted and observed air concentrations provides an indication of how well the models simulate transport, dispersion and removal processes. Hence, weaknesses in model parameterization, potentially masked in the evaluation of the models for only wet sulfur deposition could be revealed.

Seasonal and annual average air concentrations were computed from the daily samples (Table 3.6). For every season, valid 24-hour samples were available for at least half the days. Seasonally varying detection limit concentrations, determined by Barrie et al. (1984) for each site, were used when appropriate. The sulfate data at the maritime site, CO5, were corrected for sea spray by reducing the sulfate air concentrations by 0.25 of the sodium air concentration, as recommended by Barrie (personal communication, 1986).

TABLE 3.6 MEAN AIR CONCENTRATIONS OF SULFUR DIOXIDE AND SULFATE AT THE FOUR APN SITES DURING 1980.

Site	Period	Sample Size SO_2	Sample Size Sulfate	Air Concentration ($\mu g/m^3$) SO_2	Air Concentration ($\mu g/m^3$) Sulfate
ELA (C01)	Winter	86	88	2.9	1.3
	Spring	75	75	0.8	1.0
	Summer	49	52	0.6	1.0
	Autumn	66	92	2.0	1.2
	Annual	276	307	1.7	1.2
Chalk River (C03)	Winter	88	90	16.9	2.1
	Spring	86	89	6.6	2.4
	Summer	77	78	2.8	2.8
	Autumn	63	64	6.8	1.8
	Annual	314	321	8.6	2.3
Long Point (C04)	Winter	84	89	25.2	3.9
	Spring	84	91	11.6	6.0
	Summer	77	81	10.4	7.6
	Autumn	90	90	18.1	3.0
	Annual	335	351	16.5	5.1
Kejimkujik (CO5)	Winter	70	72	3.3	3.5
	Spring	70	76	1.2	3.2
	Summer	66	88	1.3	4.1
	Autumn	91	91	2.2	1.8
	Annual	297	327	2.0	3.1

4. MODEL EVALUATION APPROACH

4.1 Introduction

Model development is a continuous process in that models are refined as new theories of physical and chemical mechanisms evolve and parameterizations and input data processing techniques are refined. Model evaluations can be vehicles in this process since they can provide diagnostic information relating to the performance of a specific version of a model or specific module. Model evaluations can also assess the state of the art of modeling and identify strengths and weaknesses of the models.

A model evaluation should attempt to answer the following fundamental questions:

1) Do the model predictions concur with observations?

2) Do the models predict these results for the right reasons? (i.e., were the predictions obtained by means of credible physical and chemical parameterizations?).

To answer the first question, quality assured measurements of those predicted variables of interest are required at the same temporal scales for which simulation results from the models are applicable (i.e., for seasons and the year). An operational evaluation addresses only this question.

The second question can be best answered when continuous measurements and predictions of all model predicted variables are simultaneously available for time scales comparable to model time steps. This question is answered in a diagnostic evaluation. The detailed data are required to assess the credibility of the physical and chemical parameterizations and to compare the variability of the predictions with the variability exhibited by the measurements in response to the variations of the pertinent meteorological parameters.

Ideally, models should be evaluated via both operational and diagnostic approaches. However, this is not always possible. Since in North America many of the variables predicted by sulfur deposition models are not measured simultaneously, if at all, a thorough diagnostic evaluation of the ISDME models was not feasible. Therefore, the ISDME involved an operational evaluation accompanied by diagnostic information to identify the differences between predictions and observations.

Preceding page blank

4.2 Elements of the Model Evaluation

There is no universal set of evaluation criteria appropriate for all model evaluations. Each set needs to be tailored to at least three factors:

1) The intended applications of the model results (i.e., how the model results are to be used);

2) The spatial and temporal scales of the models; and

3) The predicted variables for which observations of sufficient quality, quantity and spatial distribution are available for a statistically meaningful model evaluation.

Regarding the first factor, the results from sulfur deposition models have been used by ecologists and policy analysts to study

1) Patterns of seasonal and annual average air concentrations and wet/dry deposition amounts;

2) Sulfur wet/dry deposition totals across states and provinces;

3) Source - receptor relationships;

4) Sulfur budgets; and

5) Sulfur fluxes across political and geographical borders.

At this time, the models cannot be evaluated for their source-receptor relationships, since a reliable means of doing so has not been devised. The models cannot be evaluated for their sulfur budgets and fluxes because of the lack of appropriate dry deposition and horizontal flux measurements.

On the other hand, models have been evaluated for spatial and temporal patterns of air concentrations of sulfur dioxide and sulfate using the 1977 and 1978 Electric Power Research Institute's Sulfur Regional Experiment (SURE) data base (Perhac, 1978). However, models have yet to be adequately evaluated for sulfur wet deposition.

Models can now be adequately evaluated for sulfur wet deposition amounts using the precipitation chemistry data available for eastern North America after 1979. Before 1980, the quantity and spatial distribution of the data rendered them inappropriate for a statistically meaningful evaluation of long-term regional models.

The ISDME models were evaluated for seasonal amounts of sulfur wet deposition according to the elements listed in Table 4.1. Seasonal periods were chosen since these temporal scales are of interest to the model user community plus seasonal evaluations are more rigorous than annual evaluations. Unlike its predecessors (see Appendix A), the ISDME compares spatial patterns of predictions to those of observations on commensurate temporal and spatial scales.

TABLE 4.1 ELEMENTS OF THE EVALUATION OF THE ISDME MODELS.

1) Replication of 1980 seasonal spatial patterns and magnitudes of sulfur wet deposition amounts across eastern North America within data uncertainty limits

2) Prediction of the magnitude and location of the maximum seasonal sulfur wet deposition amounts within data uncertainty limits

3) Replication of seasonal profiles of sulfur wet deposition amounts

4) Concurrence of aggregated seasonal and annual predictions

5) Reasonableness of predictions of seasonal average air concentrations and dry deposition amounts of sulfur dioxide and sulfate

The first two elements in Table 4.1 were selected on the basis of the usual application of the models --- the definition of seasonal and annual patterns of sulfur wet deposition amounts. The third element was selected to assess the models' capability in predicting greater wet deposition amounts for those seasons when the observed data indicate greater wet deposition occurs. The fourth element is significant if one is interested in both the seasonal and the annual results. For example, if both the seasonal and the annual effects of emission fluctuations are to be assessed, it is necessary that the model results of the four seasons add up to those obtained by an annual model execution. Otherwise, two sets of results will be obtained.

Finally, the last element was included to address, in a very limited sense, the question, were the predictions of sulfur wet deposition amounts based on reasonable predictions of other variables influencing the amounts? In other words, if a model replicates the seasonal and annual observed patterns of sulfur wet deposition amounts, how credible are the predictions of the air concentrations and dry depositions?

4.3 Model Performance Measures

4.3.1 Basis for Selection

Previous evaluations of sulfur deposition models relied primarily on the model performance measures recommended by the American Meteorological Society (AMS) (Fox, 1981). These model evaluations focused almost entirely on the comparison of statistical distributions of observations and predictions at measurement sites (e.g., means, biases, standard deviations). Spatial analyses were generally restricted to smoothed analyzed maps of data, location of maxima, spatial correlations and cross-sectional plots.

An evaluation of sulfur deposition models which focuses on the distributional statistics rather than the spatial analyses has severe limitations in that

1) The models are being evaluated on an incommensurate spatial scale (i.e., predictions, which represent averages across 10^4 km^2 areas, are compared to point measurements of a parameter that is highly variable in space);

2) Global distributional statistics (e.g., means, medians, variances, etc.) cannot reveal the spatial characteristics of the data. Hence, two completely different fields can have identical distributional properties;

3) The usual pairwise comparison of observations and predictions, in terms of average differences and/or correlations, cannot reveal the capabilities of a model to replicate spatial patterns of the observations;

4) When distributional statistics are used exclusively, those models which predict flat fields (i.e., smooth patterns with small gradients) for parameters which exhibit high spatial variations will often appear to perform well for an incorrect reason;

5) In general, the regions of poor or good agreement between the predictions and observations cannot be identified because of the widely spaced measurements.

The AMS model evaluation recommendations and the spatial analyses used in previous model evaluations alone are not sufficient to address the criteria presented in Table 4.1. Consequently, the ISDME required the development of a new approach with a greatly increased emphasis on spatial analysis. The ISDME approach uses distributional statistics from the more traditional approach plus measures relating to a robust spatial analysis (discussed in the next section) with an emphasis on the latter.

Spatial analyses can provide information not obtainable from distributional statistics, such as:

1) An assessment of model performance in regions where observations are unavailable, which is important when observations are widely spaced;

2) Specific features of the spatial distribution of the data that are important for the models to replicate (i.e., magnitude, position and shape); and

3) The degree to which the predicted spatial pattern differs from the observed pattern (i.e., by how much the predicted spatial pattern is translated or rotated and how well the predicted location and magnitudes of the gradients and maxima concur with the observed spatial pattern).

The ISDME model performance measures were selected on the basis of their:

1) Relevance to the criteria listed in Table 4.1,

2) Interpretability, and

3) Ability to objectively discriminate the behavior of the models.

These measures, listed in Table 4.2, quantify the differences between the spatial patterns (specifically magnitude, position and shape), location and magnitude of maxima and seasonal contributions of the observed and predicted amounts of sulfur wet deposition. As discussed later, values of some of these measures are calculated from predictions and observations either at ISDME evaluation sites or at interpolated grid points. Statistical significance tests are used whenever applicable. For reasons explained in Section 4.3.2 some of the measures in Table 4.2 are more reliable than others and have been designated as primary performance measures.

TABLE 4.2 THE ISDME MODEL PERFORMANCE MEASURES.

I. Replication of Spatial Pattern

  A. Magnitude

    1. Area of significant differences[*]
    2. Difference between the medians
    3. Difference between the dispersions
    4. Bias in the high deposition region
    5. Bias in the low deposition region
    6. Mean square error

  B. Position

    1. Distance between the centroids[*]
    2. Difference in the orientation of the major axes[*]
    3. Distance of the predicted pattern shift to achieve the maximum spatial correlation

  C. Shape

    1. Maximum spatial correlation resulting from predicted pattern shifting and rotation
    2. Spatial correlation without predicted pattern shifting

II. Maxima

  A. Location

    1. Distance of separation[*]

  B. Magnitude

    1. Absolute difference[*]

III. Replication of Temporal Pattern

  A. Differences in seasonal contributions to the annual total[*]

IV. Concurrence of Aggregated Seasonal and Annual Results

  A. Relative differences between the aggregated seasonal and annual predictions

---

[*] Primary model performance measures subsequently used to cluster the models according to performance.

4.3.2 Description of Measures

4.3.2.1 Spatial Patterns

The ability of the models to replicate the spatial patterns of the observations is assessed by comparing the magnitudes, positions and shapes of the observed and predicted patterns. The primary measure of *pattern magnitude* is the *percentage of area* in each of the four subregions depicted by Figure 4.1 *where the interpolated predictions are significantly different* from the interpolated observations. Interpolated values are used since neither the observations nor the predictions from most of the ISDME models are available with sufficient resolution to define the spatial patterns. The subregions are defined on the basis of the geographic regions across which the models tend to significantly under/overpredict. The unique advantage of this measure is that subregions where the models perform best and worst are readily identified by a set of quantities.

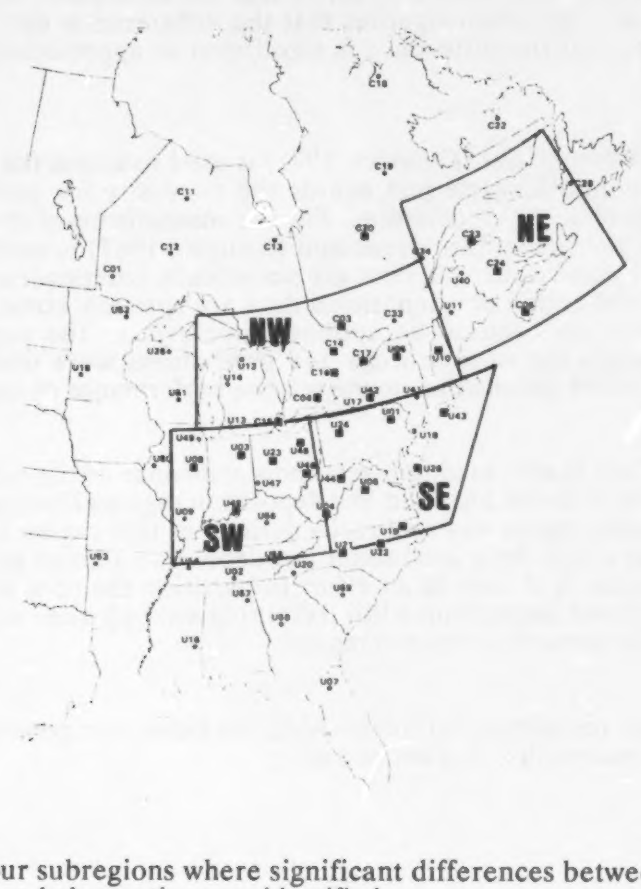


Figure 4.1 The four subregions where significant differences between interpolated predictions and observations are identified.

The percentages of subregional areas where significant differences occur are determined by examining pairs of interpolated observations and predictions at each half-degree grid cell within each subregion. Significant differences are detected when the prediction falls outside the uncertainty range about each interpolated observation, as determined via a technique discussed in Section 5. The number of cells with significant differences is determined and then divided by the number of cells in the subregion. The percentage of cells in the composite region where significant differences occur is also determined.

The next two measures of pattern magnitude relate to the distribution of the observations and predictions at the ISDME evaluation sites. The *medians* and *variances*, which are measures recommended by the AMS (Fox, 1981) are used to characterize the central tendency and dispersion of these values. Medians are better measures than means of the central tendency of the distributions because of the potential effect of outliers on the asymmetry of the distributions.

Statistical tests are normally applied to assess the significance of the differences in the medians and dispersions of the predicted and observed distributions. The "p value" associated with these tests is the probability that the differences occurred by chance. Thus, a p value of 0.05, which signifies that the difference is extremely unlikely by chance, indicates that the difference is significant at approximately the 95% confidence level.

The Mason-Whitney test (Conover, 1971) is used to assess the differences in the medians. The nonparametric test avoids the necessity for making assumptions concerning the form of the distribution. For the comparisons of the dispersion of the distributions Bartlett's test (Snedecor and Cochran, 1967) is used. However, both significance tests assume that the data are statistically independent, which is not the case here, since the sulfur wet deposition data are spatially correlated. Therefore, these significance tests cannot be applied to determine the probability that the differences between the observations and predictions were obtained by chance. However, they provide information on the relative performance of the models.

A statistical test is also used to assess the significance of the *biases* at the ISDME evaluation sites in both the high and low deposition regions illustrated by Figure 4.2. The high deposition region was arbitrarily defined as that region containing ISDME evaluation sites at which the annual sulfur wet deposition amount generally exceeds 10 kg S/ha. This value is chosen in an effort to partition the data with regards to the amount of sulfur wet deposition while retaining enough sites to characterize the distributions of the amounts in the two regions.

The bias, also recommended by the AMS, measures the general tendency of the model to under/overpredict. It is defined as

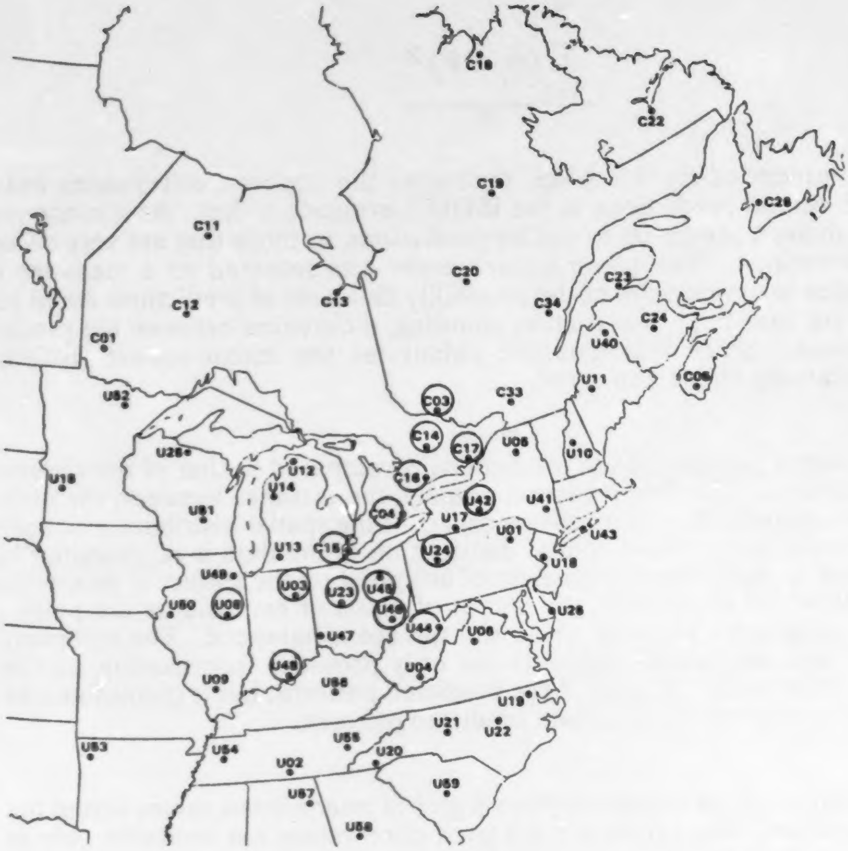


Figure 4.2 The ISDME sites in the "high deposition region" as denoted by the circles.

$$\frac{\sum_{i=1}^{N} (O_i - P_i)}{N}$$

where O_i and P_i are the observation and prediction at site i and N is the number of sites. The assessment of the statistical significance of the bias is normally based on the Student's t statistic with N-1 degrees of freedom. However, due to the spatial correlation of the data, this test cannot be used as was originally intended. Like the results of the significance tests for the median and dispersion, the results of the Student's t test are presented as a means of qualitatively comparing model performances.

The *mean square error*,

$$\frac{\sum_{i=1}^{N} (O_i - P_i)^2}{N}$$

also recommended by the AMS, measures the squared differences between the observations and predictions at the ISDME evaluation sites. As a consequence, this measure shows a sensitivity to outlier predictions, or those that are very different from the observations. The mean square error was selected as a measure of model performance in anticipation of the possibility that a set of predictions could have a very low bias, yet have large, but counterbalancing, differences between the predictions and observations. Since this measure calculates the mean square differences, no counterbalancing effects can occur.

The *pattern position* of the predictions is compared to that of the observations via three measures. The first measure, namely, the distance between the observed and predicted *centroids*, characterizes one aspect of the spatial distribution of the sulfur wet deposition amounts. If the spatial distribution of the data is represented by a three-dimensional surface, where the height of any point on the surface is proportional to the magnitude of the parameter, the center of mass, or centroid, is the point projected onto the geographic plane at which the surface is balanced. The comparison of the observed and predicted centroids not only provides information on the relative location of the center of mass of the predicted patterns, but it quantifies one aspect of the differences in the observed and predicted patterns.

The centroids are computed from a grid of interpolated values across the region of high deposition. Site values are not used since values are available only at points of measurement clustered in scattered pockets across eastern North America. The spatial distribution of the points would therefore greatly influence the location of the centroid. Since the interpolated values are available at evenly spaced grid nodes, the centroids determined from these values would be more representative of the spatial distribution of the data.

From the N pairs of interpolated observations and predictions, the centroids are computed as deposition-weighted averages of the longitude ($\Sigma\, w_i x_i$) and the latitude ($\Sigma\, w_i y_i$), where the weights are defined as

$$\frac{D_i}{\sum_{i=1}^{N} D_i}$$

and D_i is the amount of sulfur wet deposition at point i.

This subset of the modeling domain was selected to focus on the high deposition region. Consideration of the lower amounts of sulfur wet deposition across the regions to the north, south and east would have influenced the location of the centroids by virtue of their higher percentage of the total domain rather than their magnitude.

The second measure, the difference between the angles of the *major axis* of the observed and predicted patterns, quantifies the degree of concurrence in the orientation of the observed ridge of high amounts of sulfur wet deposition. From the observations and predictions at all ISDME sites but the six northernmost Canadian sites (C11, C13, C18, C19, C20 and C22), the orientation of the major axes is determined via principal component analysis. The six northern Canadian sites are not considered here since the focus of attention of this measure was the high deposition region and the immediate surroundings.

The third measure involves the distance of the predicted *pattern shift* to maximize the spatial correlation between the observed and predicted patterns. Since visual comparisons of interpolated observed and predicted patterns indicate that the predicted patterns are shifted to the south and east of the observed patterns, this measure is included to objectively quantify the apparent shift in predicted patterns.

This measure is calculated as follows: First, each of the interpolated grids of predictions is shifted in all cardinal directions in one-degree increments. After each incremental shift, the spatial correlation is computed, resulting in a grid of spatial correlations with the coordinates representing the direction and magnitude of the shift. Finally, the distances of the pattern shifts to maximize the spatial correlation are determined. The effects of pattern rotation are not considered here. Further exploration of this measure is needed before it is recommended as a model performance measure. The correlations were computed from the interpolated values rather than from the site values to compensate for local variability and errors in the measurements.

The *pattern shape* of the predictions is compared to that of the observations via the maximum *spatial correlations* resulting from pattern shifting. The spatial correlation computed before pattern shifting is also a measure to quantify the degree to which the models replicate the observed pattern shape without shifting the pattern. Both the Pearson product moment and Spearman rank correlations are computed. Since the results are very similar, only the values of the Pearson correlation coefficients are provided in this report.

When relying on the spatial correlation as a measure for pattern shape replication, pattern shifting may be necessary. The shapes of the predicted and observed patterns can be identical, yet an offset in the patterns can yield very low spatial correlations. By shifting one of the patterns an appropriate distance, the resulting correlation of 1.0 would indicate the perfect pattern shape replication.

As mentioned earlier, pattern rotation is not considered. Consequently, greater correlations would be expected for those patterns of predictions with large differences in the orientation of the major axis. Because the effects of pattern rotation are not considered, the maximum correlation is not used as a primary model performance measure.

4.3.2.2 Maxima

The differences in the location and magnitude of the *maximum* observed and predicted amounts are additional measures chosen to assess model performance for two reasons. First, the magnitude and location of the maximum deposition amounts are of interest to model users and secondly, the location and magnitude of the maximum amounts are quantifiable, within uncertainty limits.

The actual *location* of the observed maximum is uncertain because of the widely spaced monitoring sites and the unavoidable errors in the measurements and interpolations. This uncertainty must be considered in computing the distance between the observed and predicted locations of the maxima.

Two approaches, both of which consider uncertainties, were considered in determining these distances. One approach relies on spatial analyses, while the second approach relies only on the values at the ISDME sites. The first approach was more desirable because of its intrinsic advantages over the alternative approach.

First of all, since it interpolates the values between sites, this approach uses a representative analysis of the regional pattern, while the second approach considers only values at a very limited number of points that may or may not be regionally representative. Furthermore, the configuration of the monitoring network does not influence the distances calulated from the spatial analyses, since values are available at evenly spaced grid nodes.

As an example of the differences between the two approaches, if there are no monitoring sites between the site of maximum prediction and the site of maximum observation, separated by 200 km, the second approach would conclude that the model was in error by 200 km. The interpolated approach would determine a geographic domain within which the maximum observed value is likely to occur. The error of the model in predicting the location of the maximum would then be the shortest distance between the site of maximum prediction and this domain. The computed distance would not be dependent on the network configuration.

Another advantage of the interpolation approach is that it objectively quantifies the uncertainty about each grid node. Since the uncertainty assessment is based on the spatial variability of the data across the entire domain, as well as measurement errors, this approach is intrinsically capable of estimating local effects (e.g., orographical effects) on individual site values. The alternative approach assumes that the uncertainty at each site is the same.

One must be careful in applying this rigid approach since the interpolation maps could yield a predicted maximum which approximates the magnitude of a distant secondary peak value. Both values could conceivably be within the uncertainty bounds of the maximum observation, yet this approach would consider only the location of the absolute maximum value. For example, say the absolute maximum value is predicted over Ohio and a secondary maximum value is predicted over Nova Scotia, which in this case coincides with the maximum observed value. Even if the difference between the magnitudes were insignificant, this approach would calculate the distance between the Ohio and Nova Scotia locations. Given the uncertainties of the observations, it would be more appropriate to conclude that the model predicted the location of the maximum. Consequently, it is necessary to search the field for other peaks falling within the uncertainty bounds of the maximum observed value. This was done for each period of the ISDME; distances between the observed and predicted maxima were not affected.

Two approaches were also considered in quantifying the differences in the *magnitude* of the maxima, Δ_{mx}. As before, both approaches consider estimated uncertainties. The interpolation approach computes

$$\text{If } P_{mx} > O_{mx}, \quad \Delta_{mx} = \begin{cases} P_{mx} - (O_{mx} + 2\sigma_i) \text{ or} \\ 0 \text{ if } P_{mx} < (O_{mx} + 2\sigma_i) \end{cases}$$

$$\text{If } P_{mx} < O_{mx}, \quad \Delta_{mx} = \begin{cases} (O_{mx} - 2\sigma_i) - P_{mx} \text{ or} \\ 0 \text{ if } P_{mx} > (O_{mx} - 2\sigma_i) \end{cases}$$

where O_{mx} and P_{mx} are the magnitudes of the interpolated maximum observation and prediction, respectively, and the $2\sigma_i$ is the uncertainty estimate at the point of the maximum observation.

Considering data limitations, it is highly uncertain that the differences are real when the interpolated maximum prediction falls within the uncertainty envelope of the interpolated maximum observation. For these cases, the model is deemed to be on target.

The second approach computes the percentage difference between the maximum predicted site value and the maximum observed site value ±20%. The model is deemed to be on target if the predicted maximum falls within ±20% of the observed maximum.

The results of both approaches, which normally are very similar, are compared in Section 6.3. Since the uncertainty estimate of the interpolation approach is derived objectively for specific points, as compared to the subjectively determined universal uncertainty estimate of the alternate approach, the former approach is thought to be more reliable. Therefore, the interpolation approach is used to assess the performance of the models.

4.3.2.3 Temporal Patterns

Finally, since predictions should replicate the observed seasonal temporal pattern of sulfur wet deposition amounts (discussed in Chapter 3), for each model the average percentage of the annual amount of sulfur wet deposition is computed for each season. Only the data are used from those 30 ISDME evaluation sites where both observations and predictions are available for every season. The locations of these sites are denoted by the larger symbols in Figure 4.1.

4.4 Model Clustering

As discussed in Section 1.2, one of the ISDME goals is to determine which models perform best in replicating the seasonal patterns of sulfur wet deposition. A set of model performance measures has been identified to assess the performance of each model. This section discusses the approach of utilizing the measures to cluster the ISDME models according to their overall performance in replicating observed 1980 seasonal patterns of sulfur wet deposition.

There are several approaches to clustering the models according to overall performance. One approach, referred to as *model ranking*, would be to first rank the models for each performance measure. For example, the model with the lowest gross error would receive a ranking of 1 and the model with the highest gross error would receive a ranking of N, where N is the number of models. Absolute differences in magnitudes would be ignored.

The overall performance of the models would be based on the rankings for each performance measure. Those performance measures thought to be most significant could be weighted more than the others, or the performance measures could be weighted evenly, as was the case for the recent DOE/AMS evaluation of mesoscale atmospheric dispersion models (Weber and Kurzeja, 1985).

A second approach, referred to here as the *relative performance approach*, would be to first cluster the models into several categories of similar values for each performance measure. The discriminating values could be determined subjectively or on the basis of statistical significance tests. The overall performance of each model can then be determined by adding scores for each measure, where a score of 1 could be assigned to the models in the "best" category, a score of 2 to the models in the "second best" category, and so on. Performance measures with greater significance can be weighted more than other performance measures.

Finally, a third approach, referred to as the *performance standards approach*, would be to first determine whether each model passed or failed a standard for each measure. The standards would be based on modeling expectations. The overall performance of the models could be determined by the number of performance measures for which the model passes the standard. Like the other two approaches, the performance measures could be weighted.

Each approach has its advantages and disadvantages. For the model ranking approach, there are nonparametric statistical tests available to test the significance of the differences in the overall rankings. However, since the model ranking approach is strictly ordinal, it is impossible to interpret the meaning of the ranks with no other information. More specifically, the ranks are not commensurate for each performance measure. For instance, suppose N models are ranked for each of M performance measures. For Measure A, the differences between the values are very small. On the other hand, for Measure B, the range of values is very large. Given only model rankings for each measure, one would not be able to interpret the differences between the pair of models with overall rankings of 1 and N.

The disadvantage of the second approach is that the discriminating value of the performance measures are chosen subjectively. Therefore, the conclusions might not be unique. Also, in cases where there is an even distribution of the values, discrimination is difficult. If, in addition, the distribution is such that the range of values is large, one would be reluctant to assign the same score to all the models.

Another disadvantage of this approach is that one cannot distinguish whether the models in the "best" category are exhibiting acceptable behavior for any measure. As an example, consider the extreme case where all models significantly overpredict in 80% of the evaluation region. This behavior is unacceptable, yet all models would be assigned a score of 1.

Finally, the performance standards approach has the advantage that the performance of the models is assessed and can be intercompared on the basis of a known standard value. Therefore, one can relate the two groups of models to the model expectations. Furthermore, the weaknesses and strengths of each model are more easily identified. Although the standards are subjectively defined, they can reflect the expectations of the model users (e.g., NAPAP has declared that predictions of sulfur wet deposition should be expected to be within ±50% of the observations).

For the ISDME, the last approach is used to cluster the models according to overall performance. Model performance standards are prescribed for each season using the primary performance measures in Table 4.2. The other measures of Table 4.2 were not used to cluster the models because of intrinsic weaknesses or because they are in an exploratory stage.

Table 4.3 lists the criteria values for the primary performance measures. The criteria values of the last five measures in Table 4.3 pertain to maximum allowable differences between the predictions and observations. These values approximate the medians of the distribution of the models' seasonal values of each measure. Approximately half of the ensemble seasonal values are either less than or greater than these criteria values. Since the criteria values are the same for each season, it is therefore possible to compare the seasonal performances of the models as a whole. There are no criteria values for annual performance, since the overall model performance is based strictly on the seasonal performances.

The seasonal performance of the models is determined by the number of primary model performance measures passing the criteria. All but one of these measures receive a weight of one. The measure of the area of significant differences receives a weight of two or four by virtue of its assessment of

1) The significance of the differences between predictions and observations at every half-degree cell and

2) More than one attribute of the spatial pattern (i.e., it assesses differences in pattern magnitude, position and shape).

TABLE 4.3 THE CRITERIA VALUES OF THE PRIMARY MODEL PERFORMANCE MEASURES USED TO CLUSTER THE MODELS.

MEASURE	CRITERIA VALUE	SCORE
1. Area of Significant Differences	$\leq$ 20% in composite region and < 25% in each subregion	4
	OR	OR
	$\leq$ 20% in composite region and $\geq$ 25% in one subregion	2
2. Centroids	$\leq$ 75-km discrepancy	1
3. Major Axis Orientation	$\leq$ 10-degree discrepancy	1
4. Maximum Location	$\leq$ 100-km discrepancy	1
5. Maximum Magnitude	$\leq$ 30% discrepancy	1
6. Seasonal Pattern	$\leq$ 15% discrepancy	1
Maximum Score		9

The overall performance is determined by a weighted average of these seasonal scores. This set of weights reflects the observed contribution of each season on the annual sulfur wet deposition. Finally, the models are subjectively clustered into several groups according to their overall model performance scores.

5. SPATIAL INTERPOLATION

5.1 Introduction

In evaluating the performance of an air pollution model, it is essential to assess its ability to replicate validation data not only in probability distribution, but also in spatial pattern. In an effort to obtain information regarding spatial accuracy, past evaluations have utilized residuals, computed as differences between model predictions and validation data taken at monitoring sites. The analysis of residuals alone may be misleading for several reasons. First of all, field data have errors associated with them which models can not (and should not) be expected to mimic. This uncertainty should be taken into account in comparisons with model outputs. In addition, there is the problem of the "regional representativeness" of the sampling sites. A site may vary considerably from its neighbors because of a local emissions source, or the local variability associated with meteorological events. This variation is often large enough to reduce one's confidence in an individual site's ability to accurately represent a larger area.

Because of the uncertainty in individual site data, a more reasonable approach for evaluating model performance is to compare the spatial patterns derived from model outputs and from observed data. These interpolated patterns retain the general spatial features of the data by smoothing through random measurement error and local spatial variation. Associated with the estimation of spatial patterns, there is an interpolation error which results from these sources of variability. Unfortunately, most of the available interpolation procedures are undeterministic and provide no estimates of the error associated with their estimated spatial patterns. However, the statistical techniques of regression and kriging provide uncertainty estimates for interpolated surfaces. In the case of kriging, the estimates of error, as well as the smooth pattern produced, are based on the spatial covariance structure of the data. The approach makes use of weighted estimates at each interpolation point which are a function of an estimated covariance. Linear regression, on the other hand, presupposes an interpolation function whose parameters are estimated from the available data. The method produces as well a weighted estimate at each interpolated point whereby the weights are a function of its actual location as well as the estimated parameters.

One of the principal uses of kriging in the model evaluation process is to provide estimates of spatial patterns for validation and model prediction data which are then compared relative to the uncertainty in the patterns. However, this comparison does not distinguish between the components of spatial pattern, namely, *shape, positioning (orientation and location) and magnitude*.

It is important to disaggregate these spatial pattern features as much as possible in order to fully evaluate a model's performance. For example, the shape of the spatial surface produced by a model's outputs may be identical to the validation data pattern,

Preceding page blank

but the overall comparison is unfavorable due to an error in scaling. Similarly, a model may replicate the shape and magnitude of the validation data pattern, but fail in an overall pattern comparison because of a translational shift of its pattern. In order to investigate these types of model behavior, correlational measures have been utilized.

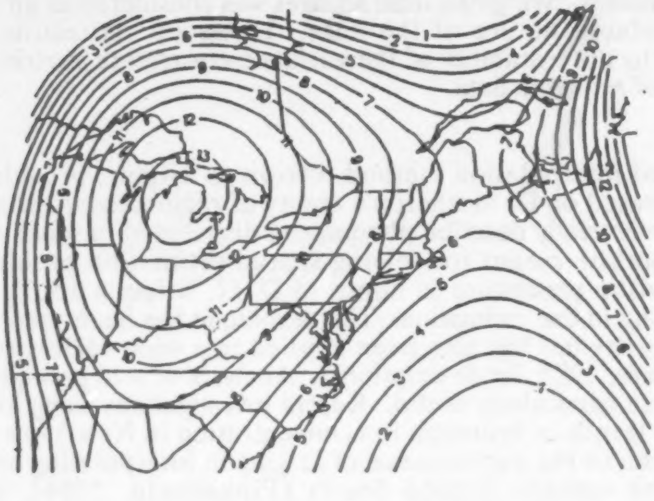
5.2 Selection of an Interpolation Method

In the analysis of a model's spatial accuracy, this evaluation concentrates on the similarity of spatial patterns which characterize validation data and model outputs. Since model outputs are relatively smooth, the patterns produced by different interpolation algorithms are similar. However, characterizations of the validation data can vary dramatically. The validation data exhibit a considerable amount of nonsystematic variation that can result in different characterizations among interpolation methods. Clearly, it is extremely important to utilize an interpolation procedure that provides the best possible characterization of the data in order to fairly evaluate the models. The procedure chosen should preserve the main characteristics of the data by smoothing through random variability. It should not be unduly influenced by single points, but should retain important spatial features. Ideally, the degree of variability in the data should influence the degree of smoothing, and some means should be available to represent the uncertainty in the interpolation estimates.

Most interpolation techniques do not have these characteristics. They are in general deterministic, and do not use data weighting schemes that utilize the degree of variability in the data. Most importantly, the user of these methods has no way to ascertain the accuracy of the characterizations that they produce. For this evaluation, we investigated two statistical techniques, standard least squares regression and a geostatistical method known as kriging.

Least squares regression was rejected for several reasons. First, it was felt that the method oversmoothed the data and lost a great deal of spatial detail (see Figure 5.1, for example). In addition, it would be difficult to justify the critical assumption of spatial independence of regression errors. Failure to meet this assumption would cast doubt on the variance estimates that were produced. In particular, a positive distance-related correlation could result in serious underestimates of variance and inappropriately narrow confidence intervals for the estimated pattern. If this occurred, models would be compared against an overly restrictive standard. The net effect would be negative assessments of model performance when model predictions were within the true uncertainty limits associated with the validation data. However, this deficiency might be offset by the enhanced residual error associated with an oversmooth regression surface. Indeed, this overestimate of interpolation variance could far overshadow the underestimates arising from spatially correlated data, thereby leading to a lenient standard. In order to better understand the magnitude of interpolation errors associated with the least squares and kriging, the interpolation variance estimates for the two methods were compared. Kriging produces lower estimates of variance in nearly all regions.

LEAST SQUARES REGRESSION



INTERPOLATION BY KRIGING

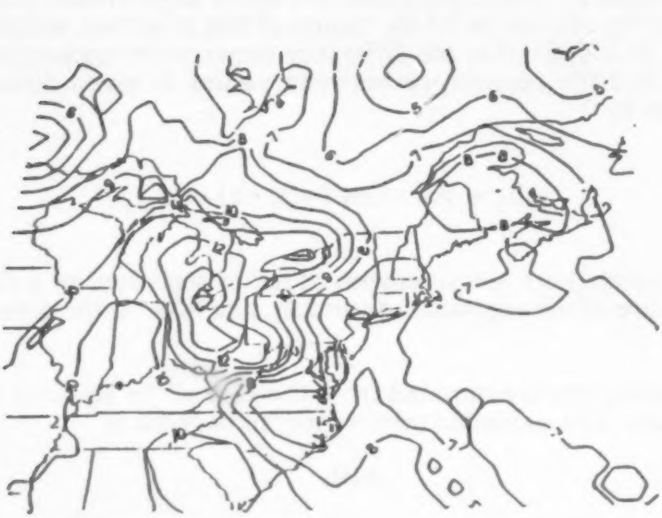


Figure 5.1 Comparison of the interpolated patterns for annual 1980 amounts of sulfur wet deposition (kg S/ha/yr) using a kriging technique and a standard least squares regression technique.

This inability of standard least squares regression methodology to sufficiently replicate the spatial surfaces is responsible for these large variance estimates. This serious drawback precluded the use of standard regression techniques as a means to estimate spatial patterns. Weighted least squares was considered as an alternative to address the correlated nature of the data. However, its results would also be questionable due to the instability of the variance-covariance matrix estimated from the small amount of available data.

A geostatistical interpolation method, known as kriging, provided an attractive alternative. It is based on G. Matheron's theory of regionalized variables, which was designed to mathematically describe phenomena distributed in space and/or time and provide an appropriate means for solving spatial estimation problems. Matheron named his estimation procedure in honor of D. H. Krige, a South African mining engineer specializing in ore estimation. This technique has been used most extensively in the mining industry but has also been applied in a wide variety of earth sciences research (see Barnes, 1980, for descriptions). Analysts of acid precipitation data have found the technique particularly useful. Recent investigations using kriging include an analysis of spatial trends in hydrogen ion concentration in New York State (Bilonick, 1983), an evaluation of the performance of kriging in interpolating acid precipitation parameters in the eastern United States (Finkelstein, 1984), and a study of temporal/spatial patterns of pH and rainfall (Eynon and Switzer, 1983).

Kriging departs from traditional deterministic interpolation methods in that the determination of interpolation weights depends on the covariance structure of the regionalized variable (in this application, wet sulfur deposition). A basic and critical part of kriging is the estimation of the model of this structure, which is known as the semi-variogram. It is defined as the difference between the underlying variance of the spatial process and the covariance between values at given distance. The semi-variogram is given by

$$\gamma(h) = \mathrm{Var} - \mathrm{Cov}\,(\, z(x_i + h) - z(x_i)\,)$$

where $z(x_i + h)$ and $z(x_i)$ are measurements at points separated by a distance of h units, "Var" is the variance of the regionalized variable, and "Cov" is the covariance.

The semi-variogram is estimated from the sum of the squared differences of all sample data values. The estimated semi-variogram is given by

$$\gamma(H) = \frac{1}{2N(H)} \sum_{i=1}^{N(H)} \left(z(x_i + H) - z(x_i) \right)^2$$

where H is a distance category dictated by the range of data; and N is the number of points in the distance category. The spatial configuration of the data locations and the interpolation site is used in combination with the semi-variogram to obtain optimal interpolation weights. The estimator that results from this procedure has several appealing characteristics. First, it is an exact interpolator: If a known point is interpolated, the exact value of the point will be reproduced. This is not true of all interpolation procedures (e.g., least squares). In statistical terms, the kriging estimator is also the best linear unbiased estimator. This implies that the estimator gives no systematic error (over or underestimation) and has the smallest estimation variance among unbiased estimators that are linear functions of the data. In addition, kriging provides an estimate of the variance between the true and interpolated values. By comparison, regression provides the best linear unbiased estimation of the parameters of the regression equation.

Kriging requires that the semi-variogram be modeled in order to estimate the weights. The semi-variogram is usually an increasing function of distance. If the variogram's value is not zero at the origin, then the intercept on the y-axis represents an estimate of the measurement error variance. In other words, it is the variance obtained from co-located sites. When kriging is applied to the predictions, the error at a co-located site is zero since a model cannot predict different depositions at the same site. On the other hand, when kriging is applied to measured depositions, it is possible to exhibit variability at the same grid site. This in fact is the case at sites such as Kejimkujik, Nova Scotia (C05).

Kriging is not a single method, but rather a family of interpolation techniques. "Simple" kriging is dependent upon the assumption that the underlying process is viewed as stationary, that is, the expected value of the process does not vary radically over the distances between observed and interpolation points. However, if the variable contains a strong trend component, simple kriging can lead to a biased estimate of interpolation variance. To redress the nonstationary problem and to account for trends, "universal" kriging has been developed. The method involves the introduction of polynomial trend terms into the system of equations used to determine optimal interpolation weights. Unfortunately, this technique relies on correctly predefining the polynomial form of trend and using a semi-variogram that does not include the trend component in its estimation. The estimation of trend-free semi-variograms has been addressed (Huijbregts, 1970; Sabourin, 1975), but has not been satisfactorily resolved. The necessity of having to preselect trend forms is also a definite drawback to universal kriging. If the wrong form is chosen, the bias in the interpolation variance may be worse than that obtained from simple kriging. For a general discussion of kriging, see Delhomme (1978) and Clark (1979).

Because of the difficulties in applying universal kriging, we have chosen to use simple kriging. As indicated above, the presence of a trend in the data may result in a biased estimate of the interpolation variance. However, in situations of moderate trend, as in the validation data, the actual interpolation estimates are not severely affected, and the seriousness of the bias in the interpolation variance is not great (Journel and Huijbregts, 1978). In addition, the direction of the bias in the interpolation variance estimates is clear. If anything, the variances produced by simple kriging will be too large. Therefore, comparisons that rely on the interpolation variance estimates will be potentially less stringent than they should be. Thus, if a model's output deviates from the confidence limits of the interpolated validation data, one can be quite certain that the deviation is significant.

A further point which needs to be mentioned is that the spatial resolution, or grid size in interpolation must be prescribed in advance. Care must be taken to ensure that it is neither too large so as to smooth the data too much, nor too small so as to produce a map with more detail than the data permit. The final choice is usually a matter of subjective judgement though it is based in part on the resolution of the validation data. The degree of smoothing in kriged analyses is, however, less dependent on grid spacing, since the weighting scheme is determined separately for each interpolated site. In this study, grid sizes of $1/4^{o}$ by $1/4^{o}$, $1/2^{o}$ by $1/2^{o}$, and 1^{o} by 1^{o} yielded very similar patterns of the validation data.

Finally, in order to ensure that kriging produces a faithful representation of the grid pattern, a comparison was made between the gridded predictions provided by TVA, RELMAP and RTM and the kriged values determined from the model predictions at only the ISDME sites. Ratios of gridded to kriged values were computed for the annual data. In general, there was excellent agreement (i.e., less than a 10% overall error) except in those regions where there were no ISDME sites (as was the case in the vicinity of Sudbury, Ontario). On the basis of this information, we deduced that kriging provides a reasonable representation of the grid pattern.

5.3 The Use of Kriging in the Model Evaluation

The primary purpose of kriging in evaluating models is to develop spatial patterns for the observed data against which model predictions could be compared. Monitoring site data were utilized to estimate the semi-variograms. Using these variograms, a surface representing these data was estimated by interpolating the data values to a $1/2^{o}$ by $1/2^{o}$ grid covering eastern North America. This region is confined to 32^{o} N - 58^{o} N latitude and 54^{o} W - 94^{o} W longitude and includes approximately 4000 grid cells.

Ideally, one would compare the predictions to interpolated observations at each grid node. Since many ISDME models did not provide gridded predictions, the ISDME site predictions provided by each model were kriged in the same manner to produce consistently derived surfaces. At each grid node, the interpolated observed and predicted values were compared, relative to the uncertainty in the interpolated values. The relative magnitude of the uncertainty across the high deposition region (depicted by Figure 4.2) was within ±50%, as shown in Figure 5.2. Measurement errors accounted for approximately one-fourth to one-third of the total uncertainty (Dennis and Seilkop, 1986). The estimates of the measurement errors, which were based on the semi-variogram fit to the observed data, are consistent with the differences observed in colocated sampling data (see Appendix C).

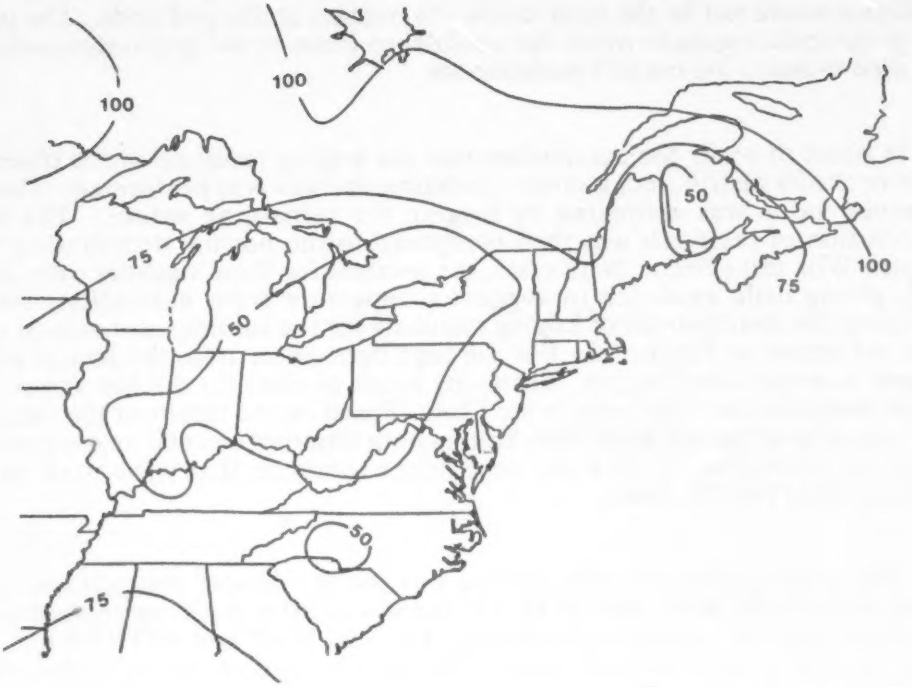


Figure 5.2 The uncertainty estimates (%) of the kriged values of annual 1980 observations of sulfur wet deposition.

A confidence interval for the difference in the observed and predicted values was constructed as follows:

$$Y_{pred} - Y_{obs} \pm 1.96 \, [\mathrm{Var}(Y_{pred}) + \mathrm{Var}(Y_{obs})]^{1/2}$$

where

Y_{pred} = kriging estimate for the model prediction at the grid node

Y_{obs} = kriging estimate for the observed data at the grid node

Var = kriging variance estimate.

If the deviations of the kriged values from the true values that they estimate have a normal (Gaussian) distribution, this is a 95% confidence interval. Failure of the confidence interval to include zero can be viewed as strong evidence that the model prediction would not be the same as the observation at this grid node. The percentage of a geographic region in which the confidence limits at the grid nodes excluded zero was used to assess the model's performance.

In order to verify the assumption that the kriging residuals of the observed data were normally distributed, a cross-validation analysis was performed. That is, each observed value was estimated by kriging the remaining values. The empirical distribution of residuals was then compared to the normal distribution using the Shapiro-Wilk test (1965). In all cases, the p-values for these statistics were larger than 0.25, giving little evidence to suggest a departure from normality. Histograms depicting the distributions of kriging residuals for the summer and annual validation data are shown in Figure 5.3. For the high deposition area, the histograms clearly suggest a normal distribution. Normality is not as clear for the low deposition area, where interpolation "edge" effects are likely. Based on the results of the statistical tests for normality of the residuals from kriging both observations and model predictions, it is not unreasonable to view the confidence intervals that were used as being at approximately the 95% level.

The same procedure was carried out for the model predictions. With the exception of RELMAP and ASTRAP, there was little evidence to suggest that the residuals were not normally distributed. For RELMAP and ASTRAP the departure from normality was relatively minor. The distributions of the residuals arising from kriging these model predictions were symmetric but longer tailed than expected in a normal distribution.

A problem which could arise from kriging is associated with the fact that some models do not produce "smooth" patterns of predictions. For these cases, the proposed confidence interval will be inflated in size because of disproportionately large interpolation errors for the model predictions. Therefore, the more variable model prediction pattern will be less likely to significantly deviate from the observed pattern. The magnitude of acceptable model deviations in this case is not only influenced by the uncertainties of the observations, but also by the uncertainties in the estimates of the model predictions. This clearly represents an unfair advantage over models with smooth predictions.

According to Table 5.1, interpolation errors in characterizing annual prediction grid values for most models did not exceed 25% of the interpolation errors associated with estimating the observation pattern. The magnitude of the interpolation errors for the annual RELMAP predictions were similar to, while those for ASTRAP were approximately 50% of, the magnitude of the interpolation errors for the annual observations. Judging from the results discussed in Section 6, the large annual interpolation errors for RELMAP could be attributed to two seasons - summer and spring. On the other hand, the relatively large annual interpolation errors of ASTRAP could be attributed to each season. Consequently, the results of the spatial analysis of RELMAP predictions in summer and spring and those for ASTRAP for each season were interpreted in light of the inflated size of the confidence interval.

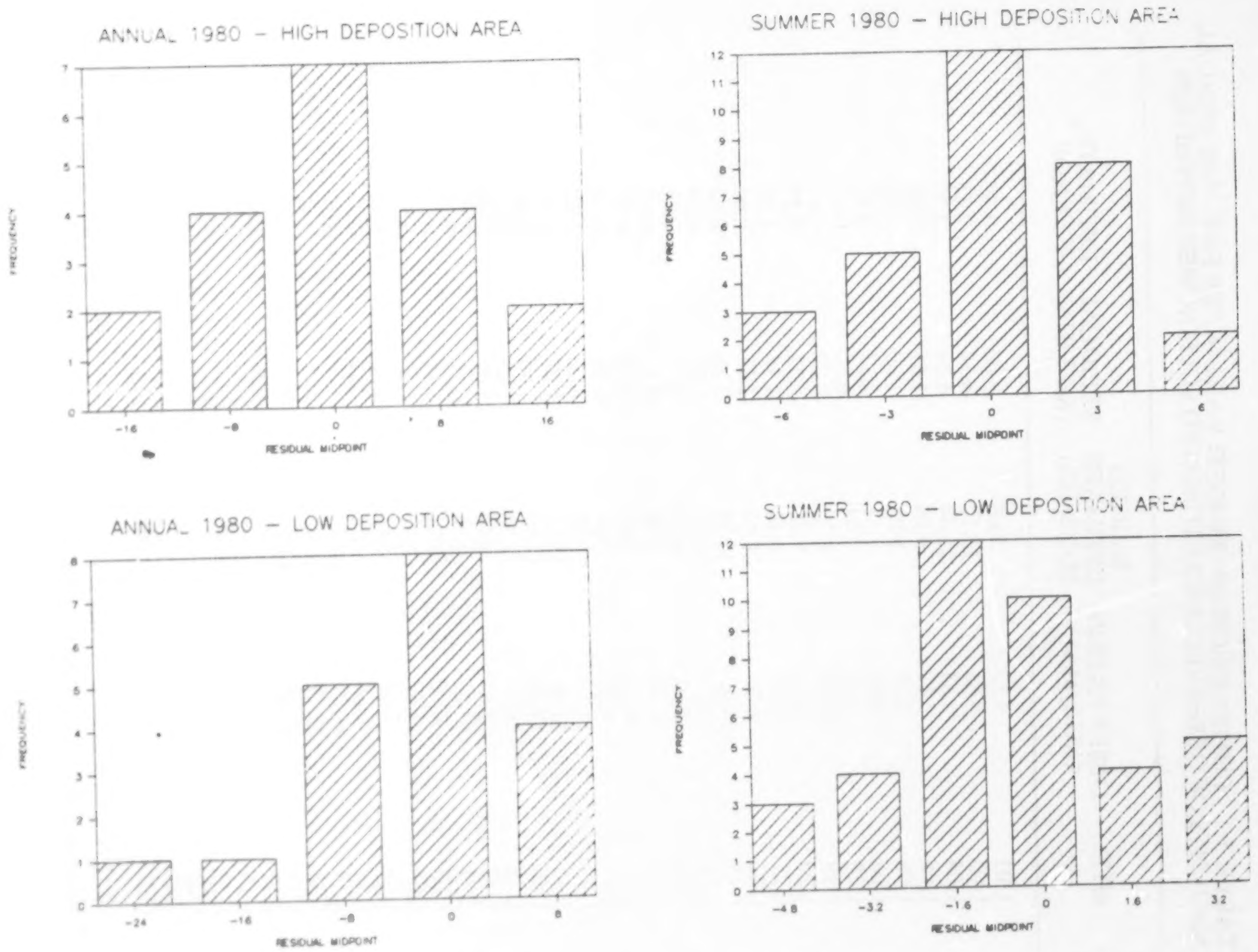


Figure 5.3 The histograms of kriged residuals for the summer and annual 1980 amounts of sulfur wet deposition (kg S/ha).

TABLE 5.1 COMPARISON OF KRIGED VARIANCES FOR 1980 ANNUAL AMOUNTS OF OBSERVED AND ESTIMATED SULFUR WET DEPOSITION.

SITE	OBSERVATION (KG S/HA)	KRIGED ESTIMATE (KG S/HA)	RESIDUAL (KG S/HA)	ESTIMATE OF ERROR*
U01	5.57	7.65	-2.08	-1.39
U03	8.63	13.17	-4.54	-2.23
U04	5.33	11.34	-6.01	-3.37
U05	7.47	9.08	-1.61	-0.86
U06	7.03	8.83	-1.80	-1.03
U08	9.93	7.81	2.12	0.87
U10	7.13	6.98	0.15	0.07
U19	6.23	6.25	-0.02	-0.01
U20	8.53	9.25	-0.72	-0.27
U21	10.07	6.51	3.56	1.85
U22	6.23	7.48	-1.25	-0.64
U23	14.40	11.24	3.16	2.22
U24	9.25	10.25	-1.00	-0.54
U26	6.93	6.23	0.70	0.31
U41	6.27	6.58	-0.31	-0.17
U42	8.33	8.42	-0.10	-0.07
U43	6.23	5.97	0.26	0.12
U44	14.53	8.53	6.00	3.62
U45	11.00	15.12	-1.00	-4.31
U46	13.47	11.99	1.48	0.92
U48	8.67	9.70	-1.03	-0.41
U52	2.67	3.70	-1.03	-0.40
C01	2.05	2.38	-0.33	-0.12
C03	9.00	10.31	-1.31	-0.67
C04	11.67	12.67	-1.00	-0.52
C05	5.83	8.45	-2.62	-1.03
C13	2.80	4.62	-1.82	-0.57
C14	11.03	10.92	0.11	0.06
C15	15.77	11.00	4.77	2.61
C16	12.20	11.70	0.50	0.29
C17	13.50	9.62	3.88	2.38
C19	2.77	4.14	-1.37	-0.44
C20	5.13	5.60	-0.47	-0.18
C22	3.53	4.80	-1.27	-0.29
C23	9.20	7.33	1.87	0.77
C24	9.13	7.44	1.69	0.85

---

* Determined by dividing the residual by the standard deviation at that site.

6. MODEL EVALUATION RESULTS

6.1 Introduction

With few exceptions, each of the eleven ISDME models predicted seasonal and annual totals of sulfur wet and dry depositions and seasonal and annual averages of sulfur dioxide and sulfate air concentrations at most of the 66 ISDME sites depicted in Figure 3.9. Table 6.1 summarizes the 1980 seasonal and annual predictions provided by each model. Since some of the northern Canadian sites were outside their domain, not all models provided predictions at all 66 sites. UMACID predicted air concentrations and deposition amounts at 47 sites, most of which are included in the ISDME evaluation data set. All models provided predictions at the fifteen sites in the high deposition region (Figure 4.2).

TABLE 6.1 SUMMARY OF THE PREDICTIONS PROVIDED BY EACH MODEL.

MODEL	NUMBER OF SITES	SO_2/SULFATE WET DEPOSITION	SO_2/SULFATE DRY DEPOSITION	SO_2/SULFATE AIR CONCENTRATION
AES	65	both	both	both
ASTRAP	66	bulk S	both	both
MOE	66	bulk S	bulk S	both
RELMAP	62	both	both	both
RIVAD	65	both	both	both
RTM-II	60	both	both	both
SERTAD	65	both	SO_2 only	both
SIMPMOD	64	both	both	both
STATMOD*	64	both	both	both
TVA	66	both	both	both
UMACID	47	both	both	both

* Annual results are obtained from a separate model run rather than summing or averaging seasonal results.

Appendix D contains three sets of tables consisting of the site values and distributional statistics (see footnote). The seasonal and annual site predictions and observations of sulfur wet deposition are provided in Tables D.1-D.5. Hereafter, observations are those amounts of sulfur wet deposition passing the screening criteria as discussed in Chapter 3. Standardized residuals (i.e., residuals divided by the observations) at each of the ISDME evaluation sites are provided in Tables D.6-D.10. Finally, Tables D.11-D.15 compare the distributional statistics of the predictions and observations in the low and high deposition regions as well as the entire region.

The remainder of Chapter 6 presents and discusses the model evaluation results for each season of 1980. Since overall model performance is based on only the seasonal evaluation results, the annual 1980 results are presented in Appendix E. The site predictions and observations and the spatial patterns of the predictions and observations of sulfur wet deposition amounts are compared for each season and the year of 1980 via the model performance measures discussed in Chapter 4.

6.2 Seasonal Temporal Patterns

The observed amounts of sulfur wet deposition exhibited a noticeable seasonality as a result of the seasonal variations in pertinent physical and chemical processes (e.g., sulfate scavenging and sulfur dioxide transformation rates). This seasonality aspect should not be ignored when evaluating sulfur wet deposition models, even if they are to be applied for annual periods. A model should replicate the observed temporal pattern.

The temporal patterns of the predictions and observations were determined by averaging the sulfur wet deposition amounts at the 30 ISDME sites where both predictions and observations were available for each model and for every season. The relative contribution of each season to the annual total was computed for the set of observations and predictions (see Table D.16). The mean seasonal deposition amounts predicted by each model were compared with the mean seasonal observed amounts to qualitatively ascertain whether any of the ISDME models failed to reasonably replicate the seasonal temporal pattern (Figure 6.1).

The seasonal pattern of the observations, represented by the bold lines in the graph, indicated that approximately two-thirds of the annual amount of sulfur wet deposition occurred in the summer (38%) and the spring (30%). The winter and autumn amounts accounted for approximately 16% each.

Footnote: Just prior to publication, the authors were notified of a SERTAD programming error which understated the seasonal and annual values of sulfur wet deposition at site U42, near central New York. This error explained the artificially strong gradients apparent in the winter, spring and summer spatial patterns across the New York area.

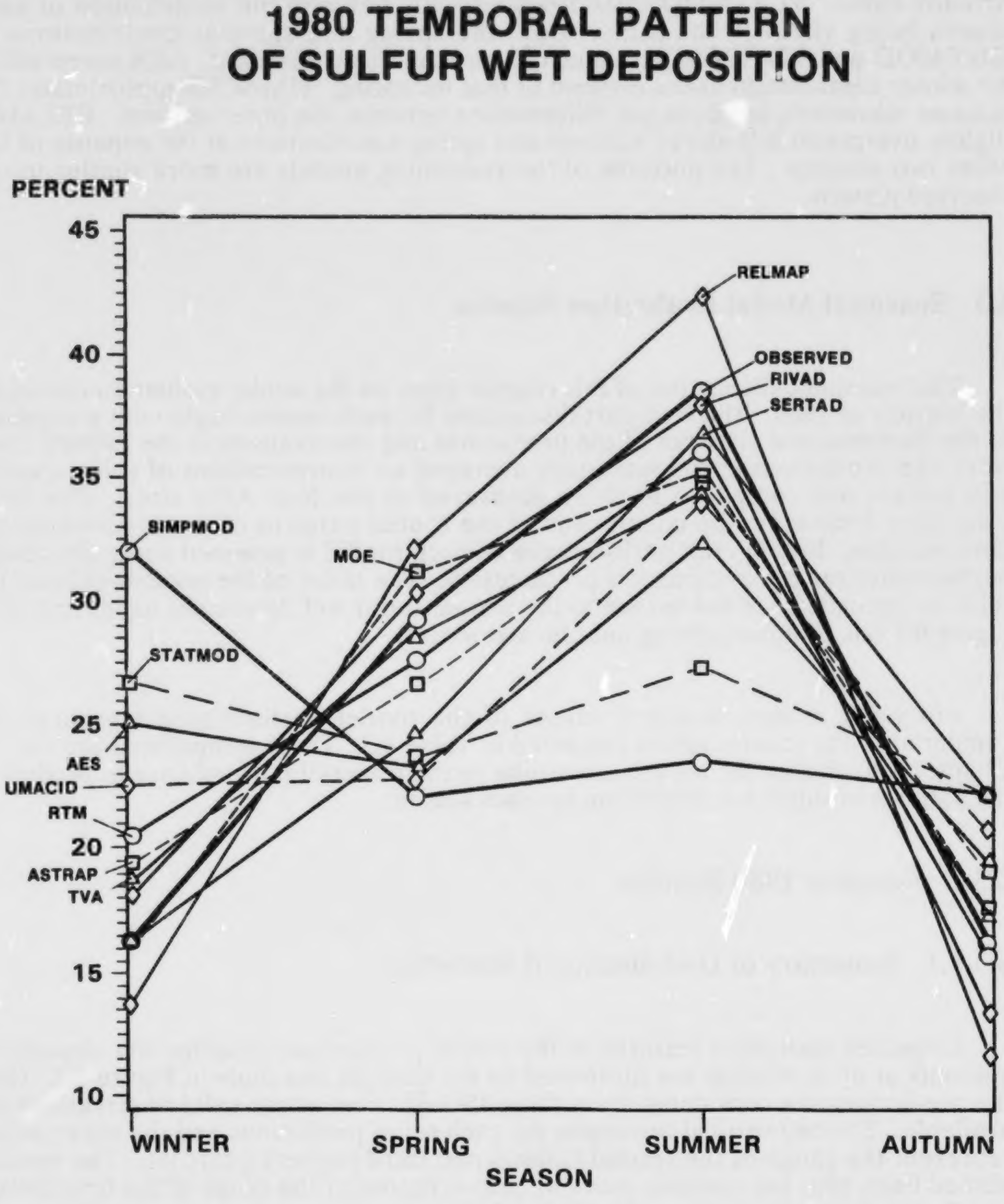


Figure 6.1 The mean percentage of annual 1980 sulfur wet deposition observed and predicted for each season at 30 ISDME sites.

Several models poorly replicate the observed seasonal pattern. SIMPMOD winter predictions account for nearly one-third of its annual prediction, more than double the observed contribution. Meanwhile, its spring, summer and autumn contributions are virtually equal. STATMOD's pattern is virtually flat with the contribution of each season being virtually the same. Both the spring and summer contributions of SIMPMOD and STATMOD are much lower than those observed. AES overpredicts the winter contribution at the expense of that for spring. UMACID approximates the summer maximum, but does not differentiate between the other seasons. RELMAP slightly overpredicts both the summer and spring contributions at the expense of the other two seasons. The patterns of the remaining models are more similar to the observed pattern.

6.3 Seasonal Model Evaluation Results

The remaining discussions of this chapter focus on the model evaluation results for the seasons of 1980. The four-part discussions for each season begin with a summary of the distributional statistics of the predictions and observations at the ISDME sites. Next, the predictions of the seasonally averaged air concentrations of sulfur dioxide and sulfate are compared to those measured at the four APN sites. The third discussion focuses on the comparison of the spatial patterns of the predictions and observations. Finally, the performance of each model is assessed using the model performance measures discussed in Chapter 4. The order of the seasons reflects the relative importance of the season to the annual sulfur wet deposition as illustrated in Figure 6.1 (i.e., summer, spring, autumn and winter).

For each model, seasonal values of the model performance measures are compared to the criteria values presented in Table 4.3. These comparisons are used in Chapter 7 to cluster the models according to their overall performance in predicting the patterns of sulfur wet deposition for each season.

6.3.1 Summer 1980 Results

6.3.1.1 Summary of Distributional Statistics

Unpaired statistical features of the model predictions of sulfur wet deposition amounts at up to 45 sites are illustrated by the quartile box plots in Figure 6.2. Only the predictions are considered from those ISDME sites where valid observations are available. The two vertical rectangles for each set of predictions and the observations represent the range of the second (lower) and third (upper) quartiles. The vertical dashed lines, plus any symbols above or below, represent the range of the first (lower) and fourth (upper) quartiles. The "o" and "*" symbols represent values which would occur about once in 20 ("o") and once in 200 times ("*") if the distribution were normal. The mean value is indicated by "+", while the median is represented by "*---*". Standard deviations of the distributions are provided to the left of the first quartiles.

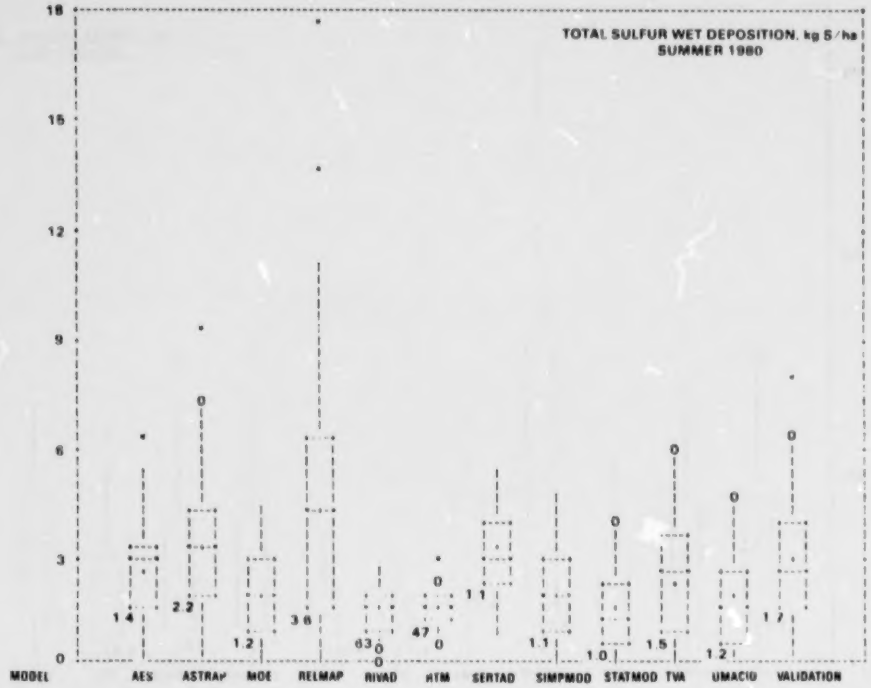


Figure 6.2 **Summer 1980** box plots of the predictions and observations of sulfur wet deposition (kg S/ha).

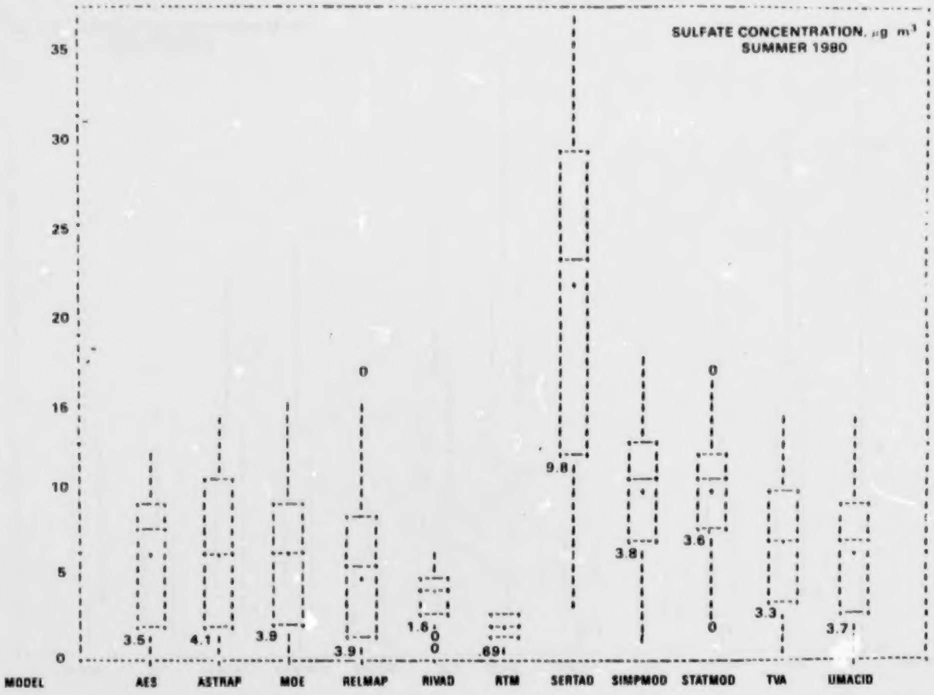


Figure 6.3 **Summer 1980** box plots of the predictions of sulfate air concentrations ($\mu g/m^3$).

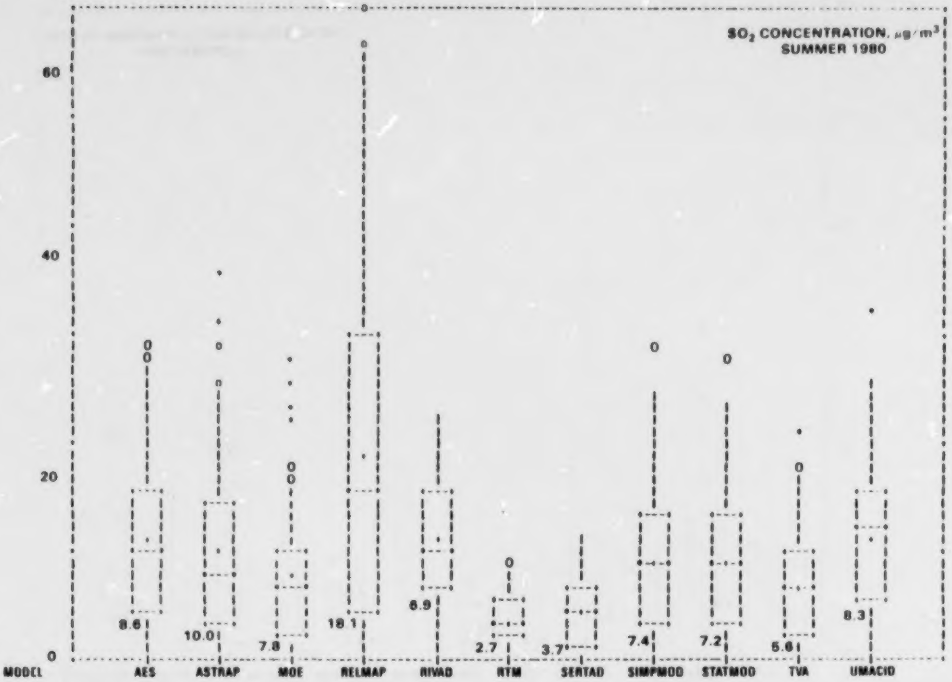


Figure 6.4 Summer 1980 box plots of predicted sulfur dioxide air concentrations (µg/m³).

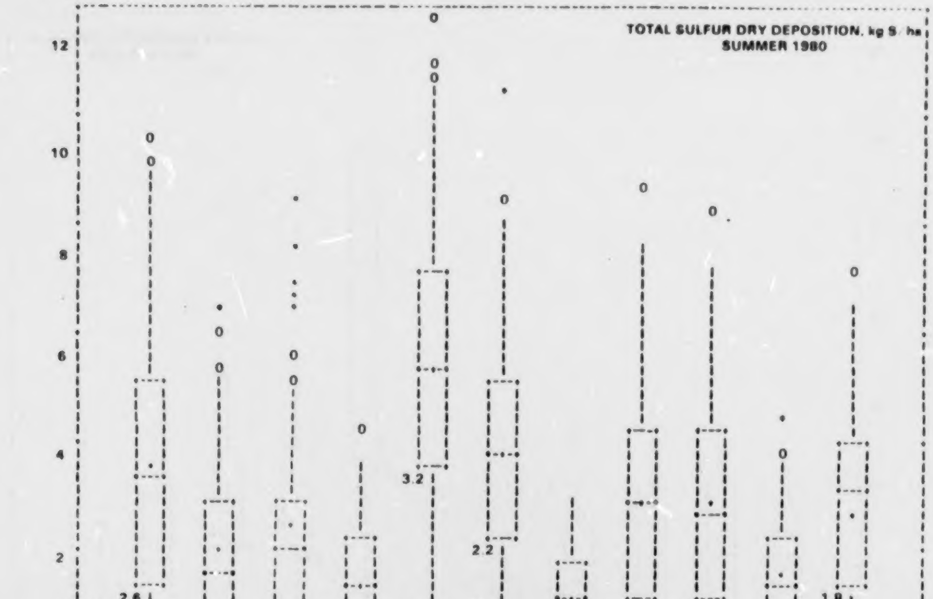


Figure 6.5 **Summer 1980** box plots of predicted sulfur dry deposition (kg S/ha).

This figure indicates that for most models, the distribution of sulfur wet deposition predictions resembles that of the observations. However, both RIVAD and RTM substantially underpredict and their standard deviations are much lower than that of the observations. On the other hand, RELMAP tends to substantially overpredict and its standard deviation is twice that of the observations.

Figures 6.3-6.5, which are quartile box plots of the predicted summer average air concentrations of sulfate and sulfur dioxide and sulfur dry deposition amounts at all ISDME sites where predictions are available, illustrate the behavior of the models. This information is of limited use since sufficient measurements of these quantities are not available. However, the figures provide the means to

1) Intercompare model results for which there are no measurements,

2) Assess the magnitude of the model process rates, and

3) Assist in explaining model performance.

Figure 6.3 illustrates the very high SERTAD sulfate air concentrations, while Figure 6.4 shows that SERTAD sulfur dioxide air concentrations are considerably lower than those of the other models. These differences can be explained by

1) The very high SERTAD conversion of sulfur dioxide to sulfate near the sources (40% of the total mass of sulfur dioxide is converted to sulfate)

2) The lack of consideration of sulfate dry deposition.

The fact that RIVAD and RTM underpredict the amounts of sulfur wet deposition (Figure 6.2) can be explained by the tendency of both models to predict low sulfate air concentrations (Figure 6.3) and for RTM low sulfur dioxide air concentrations as well. These low predictions are attributed in part to the relatively high predictions of sulfur *dry* deposition (Figure 6.5), which tend to be factors of two to four higher than their predictions of sulfur *wet* deposition (Figure 6.3).

Finally, the relatively high RELMAP sulfur dioxide air concentrations of the third and fourth quartiles (Figure 6.4) prompted an investigation whichdiscovered that all the emissions were inadvertently injected into the surface layer, rather than each of the three layers. An additional model application for a one-month period indicated that the injection of all emissions into the surface layer did not have a substantial effect on RELMAP predictions of sulfur wet deposition, but did cause higher air concentrations, particularly for sulfur dioxide.

6.3.1.2 Comparison of Predictions and APN Measurements

During 1980 daily average air concentrations of sulfur dioxide and sulfate were measured at four Air and Precipitation Monitoring Network (APN) sites in Canada. These sites, where sulfur wet deposition was also monitored, are located in western Ontario (C01), southern Quebec (C03), southern Ontario (C04) and southern Nova Scotia (C05) (see Figure 3.9). Seasonal and annual means of these daily concentrations were computed and compared to the corresponding means predicted by the models at each site. Since data are available from only four sites, the general tendencies of the models deduced from these figures are not and should not be extrapolated to other sites. The comparisons are used here to identify outlier models at these sites only.

Figure 6.6 compares the predicted and observed air concentrations of sulfate and sulfur dioxide for the summer at these four sites. Three of the four statistical models, SERTAD, SIMPMOD and STATMOD, substantially overpredict sulfate air concentrations. The SERTAD predictions are as much as six times greater than the observations. Similarly, Figure 6.3 indicates that the means and standard deviations of the SERTAD sulfate air concentrations are more than twice those of any other model.

RTM tends to underpredict by a factor of two the sulfate air concentrations at the APN sites. Since Figure 6.3 also indicates that the RTM predictions are much lower than those of other models, it is likely that this model substantially underpredicts the sulfate air concentrations elsewhere.

For sulfur dioxide air concentrations, all of the models, with the exception of RTM and SERTAD, tend to overpredict at the APN sites. RIVAD tends to overpredict by as much as a factor of nine. However, the distribution of its predictions at the ISDME sites (Figure 6.4) is similar to that of most other models.

6.3.1.3 Spatial Analyses

The comparisons of the kriged spatial analyses of predictions and observations in Figure 6.7 enable one to assess the ability of the models to replicate observed patterns. In the upper left corner of each page in this series of plots is the kriged analysis of the summer observations. Highlighted are the isopleths that are multiples of 4 kg S/ha.

The kriged analysis of the observations shows that the greatest amounts of sulfur wet deposition are found across Ohio and western Lake Erie. A ridge of large amounts is apparent from Illinois and Wisconsin to Ohio and continuing northeastward to the Maritime provinces. Several secondary peaks are apparent to the north and northeast of Lake Ontario and in the vicinity of southern Lake Michigan. Because of the uncertainties of the measurements, however, the actual existence of all of these peaks is uncertain. Also apparent is the region of minimun amounts along the East Coast of the United States, which corresponds to the region of minimum summer

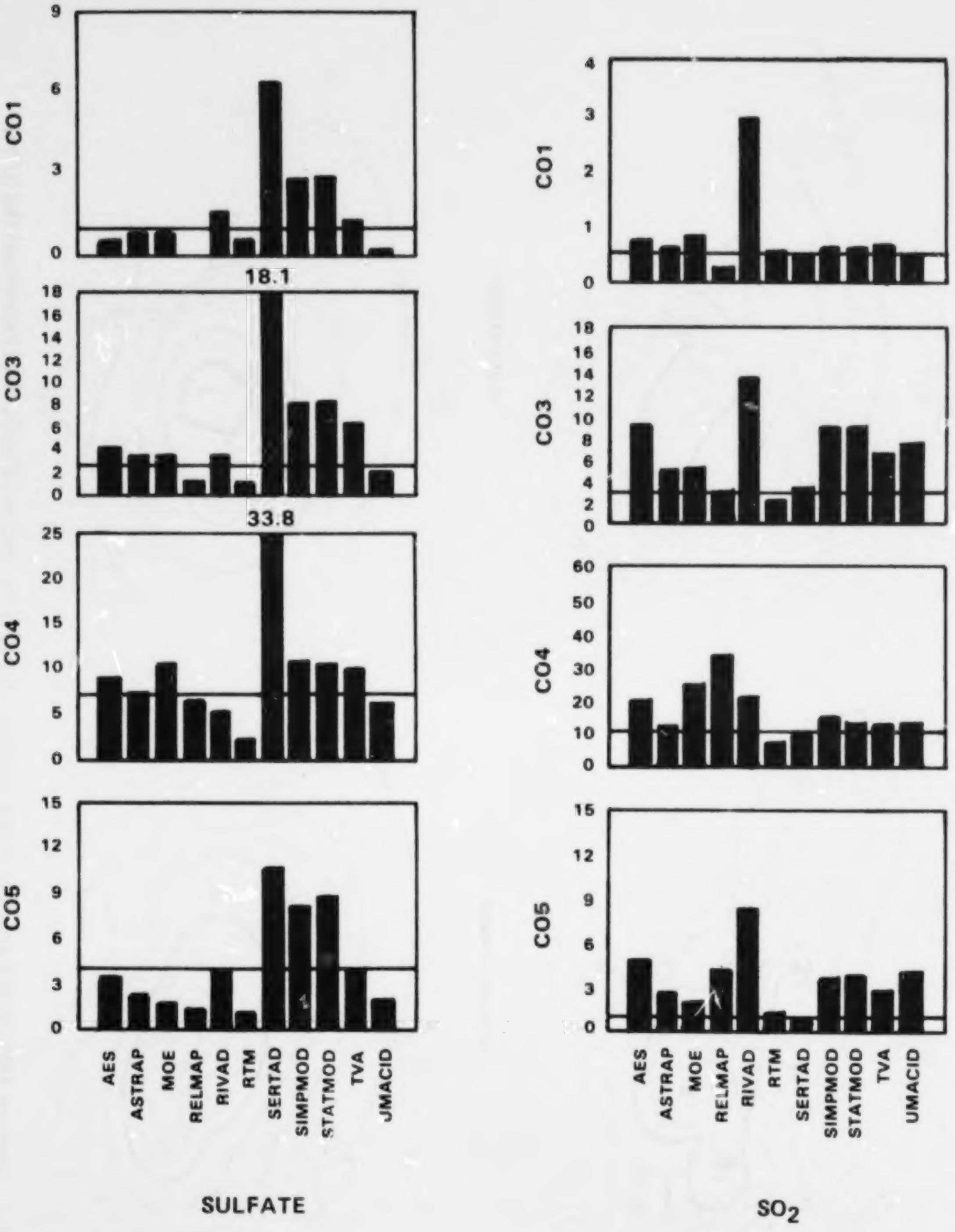


Figure 6.6 Comparison of the **summer 1980** observations and predictions of the air concentrations ($\mu g/m^3$) of sulfur dioxide and sulfate at the APN sites.

Reproduced from
best available copy.

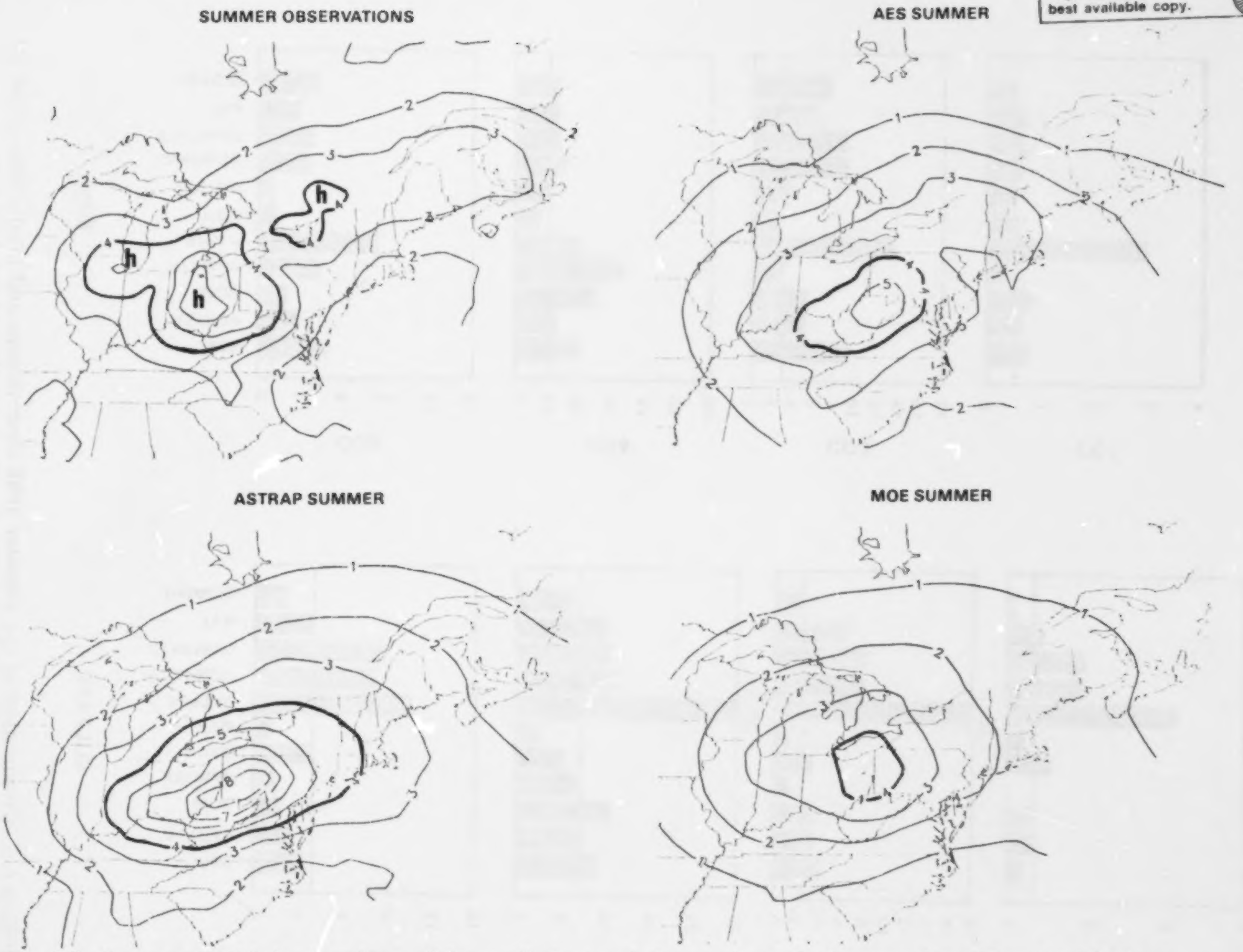


Figure 6.7 **Summer 1980** kriged analyses of the predictions and observations of sulfur wet deposition (kg S/ha).

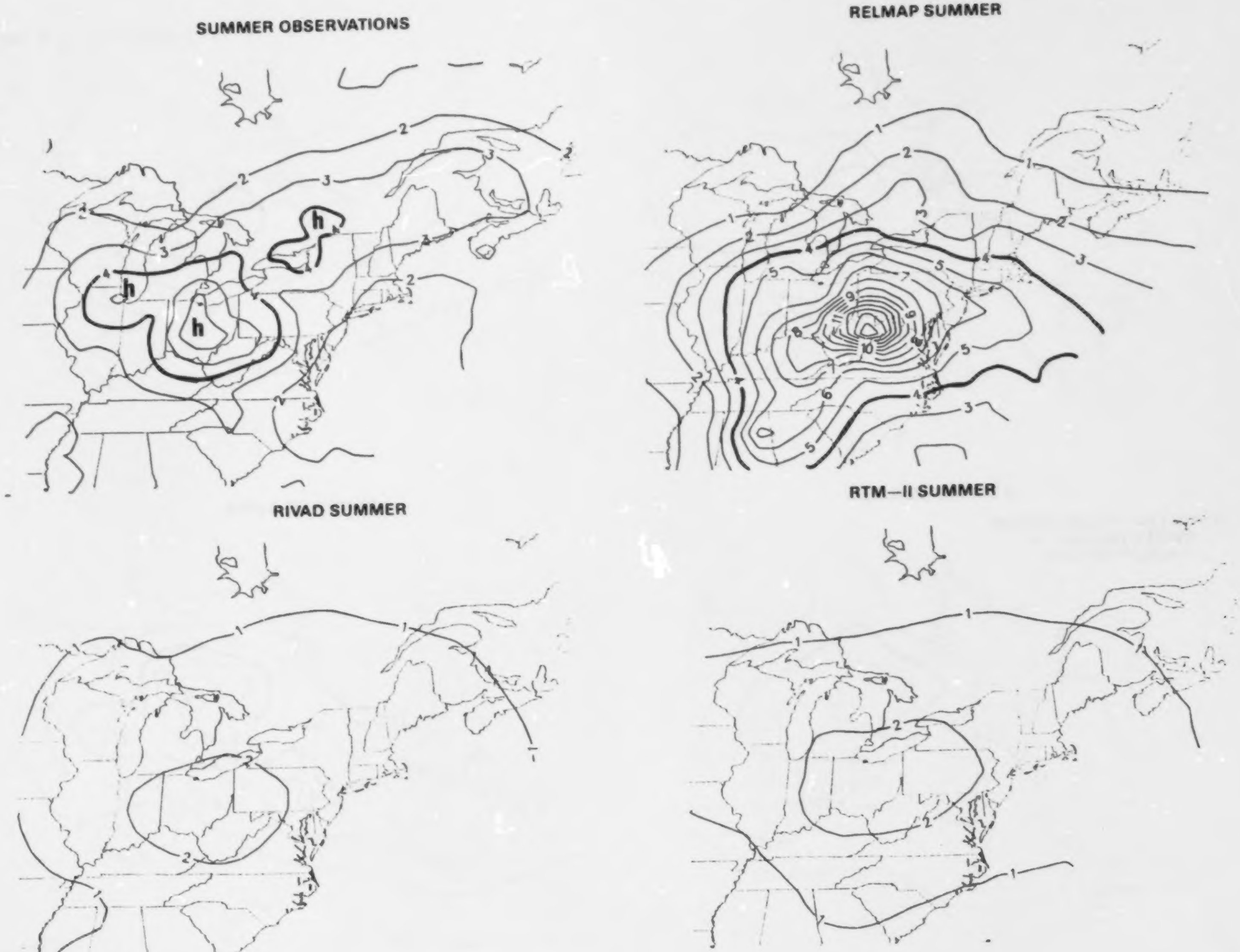
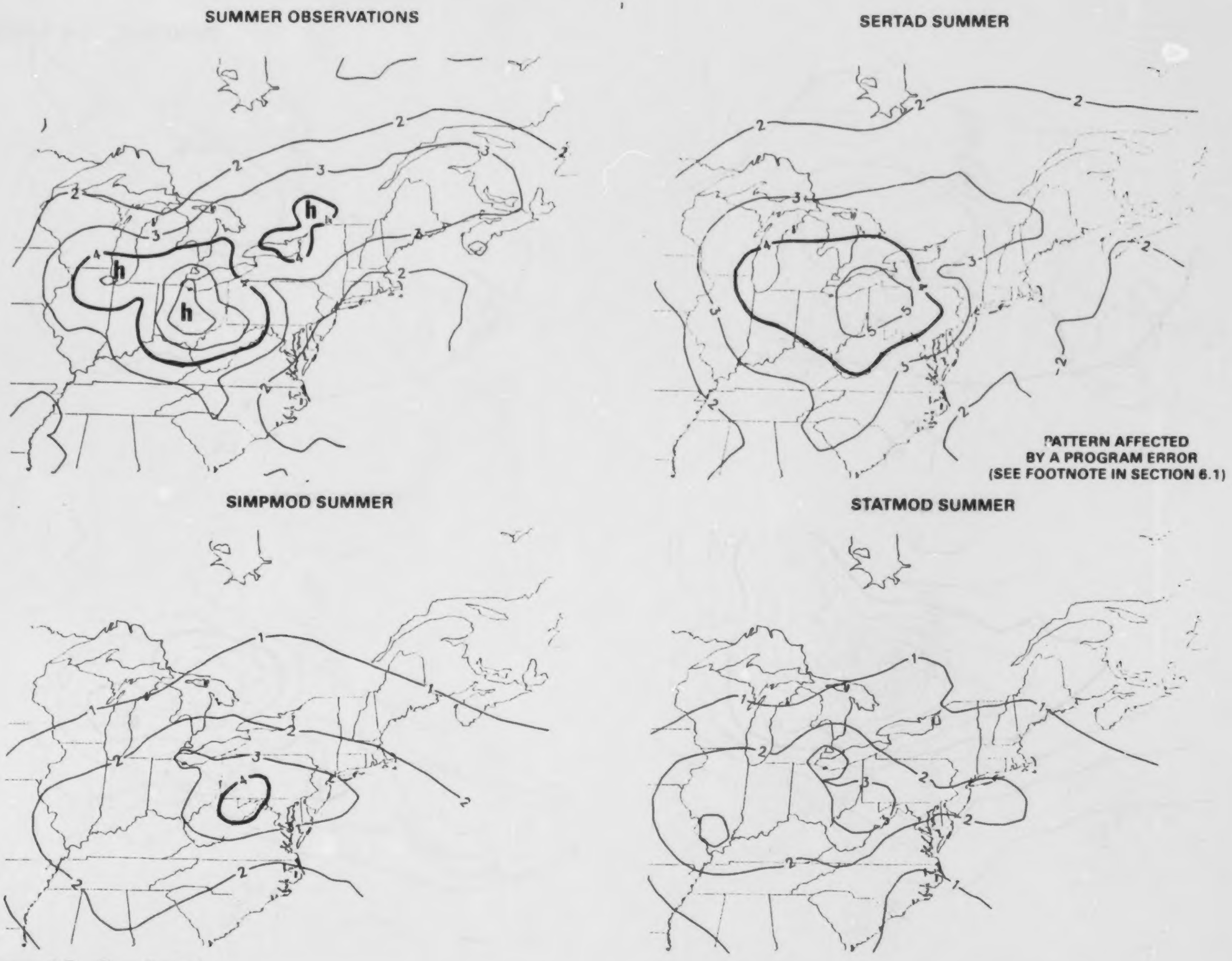


Figure 6.7 Continued



SUMMER OBSERVATIONS

SERTAD SUMMER

PATTERN AFFECTED BY A PROGRAM ERROR (SEE FOOTNOTE IN SECTION 6.1)

SIMPMOD SUMMER

STATMOD SUMMER

Figure 6.7 Continued

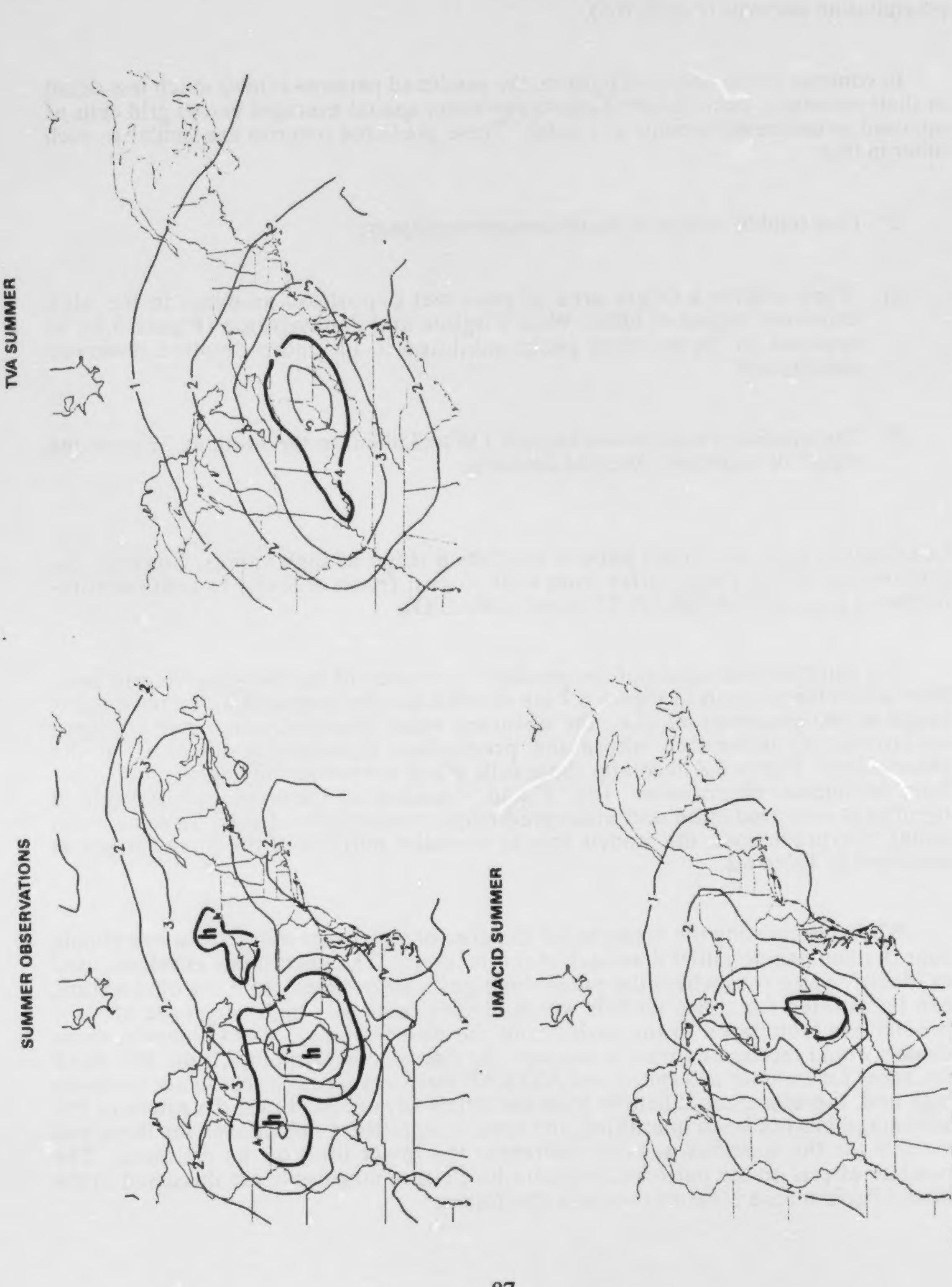


Figure 6.7 Continued

precipitation amounts (Figure B.5).

In contrast to the observed pattern, the predicted patterns exhibit much less detail in their structure, since the predictions represent spatial averages across grid cells as opposed to the measurements at a point. These predicted patterns are similar to each other in that

1) They tend to consist of nearly concentric ellipses;

2) They exhibit a single area of peak wet deposition amounts in the high emissions region of Ohio, West Virginia and Pennsylvania (Figure 3.2), as opposed to the multiple peaks exhibited in the more detailed observed pattern; and

3) The predicted maxima are located 130 to 350 km to the south and east of the region of maximum observed amounts.

In addition, each predicted pattern exhibits a ridge of high values, however the orientation of this ridge varies from west-to-east (most models) to southwest-to-northeast (e.g., AES, ASTRAP, TVA and UMACID).

The interpolated values of the predictions at each of the half-degree grid cells, from which the patterns in Figure 6.7 are derived, can be compared to the uncertainty range of the observations (i.e., the observed value plus and minus two standard deviations) to determine where the predictions significantly differ from the observations. Figure 6.8 illustrates those cells where the predictions significantly differ from the summer observations. The "+" and "-" symbols on the maps represent cells of significant overprediction and underprediction, respectively. Based on patterns of under/overpredictions, the models appear to cluster into one of the three categories described in Table 6.2.

When interpreting the measure for the area of significant differences, one should bear in mind one potential drawback of this measure: The uncertainty envelope, used to identify those cells where the prediction significantly differs from the observation, can be inflated for those models which exhibit relatively high variances of their predictions (compared to the variance of the observations). Consequently, these models would receive an unfair advantage. As discussed in Sections 5.1, the RELMAP variances for summer and spring and ASTRAP variances for each season are relatively high and, therefore, could benefit from the unfair advantage. Since the extent of this advantage has not been quantified, the area of significant differences for these two models for the specified seasons represent the lower limit of the estimate. The possible effects on the performance score for pattern magnitude are discussed in the Model Performance Measures sections that follow.

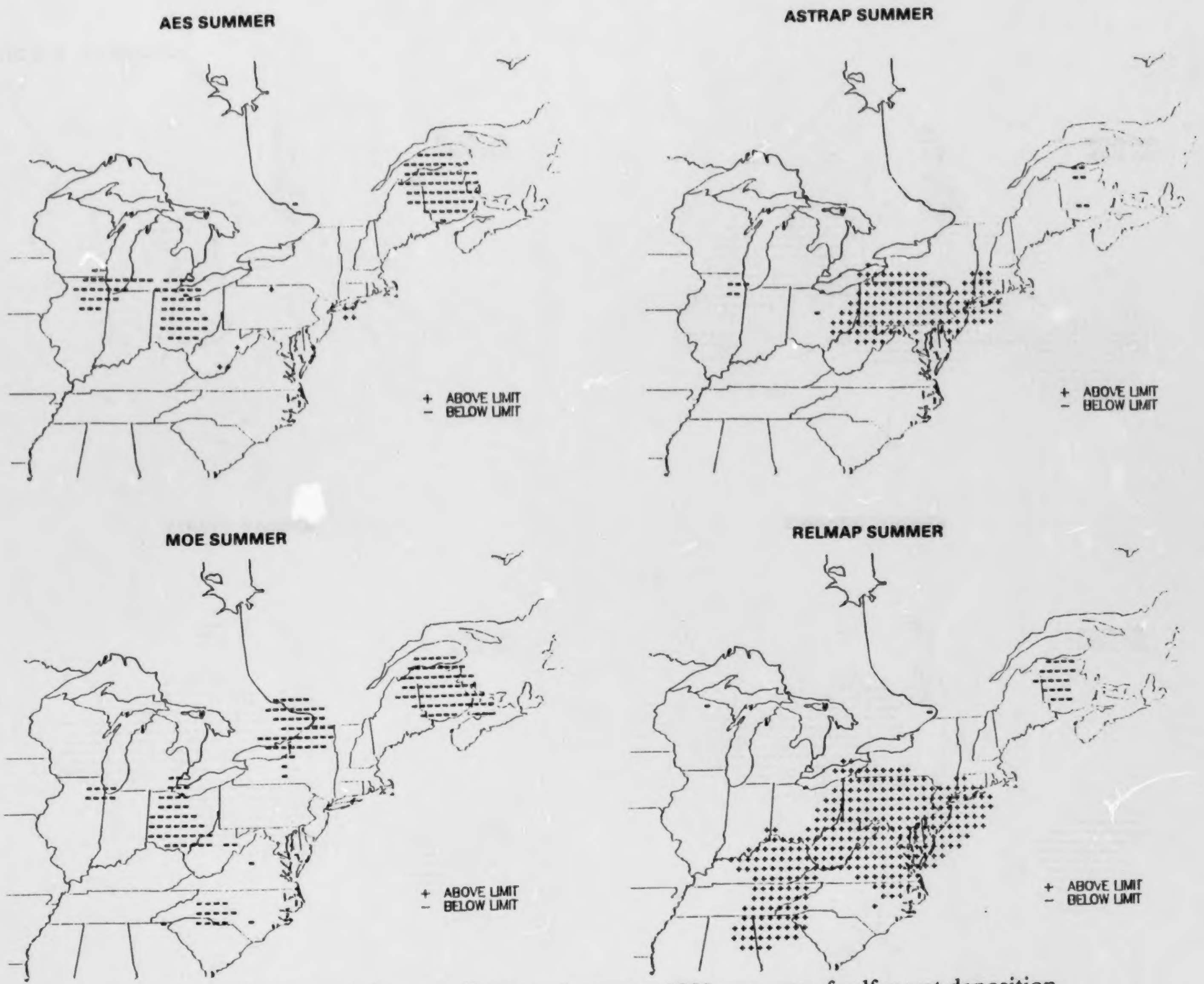
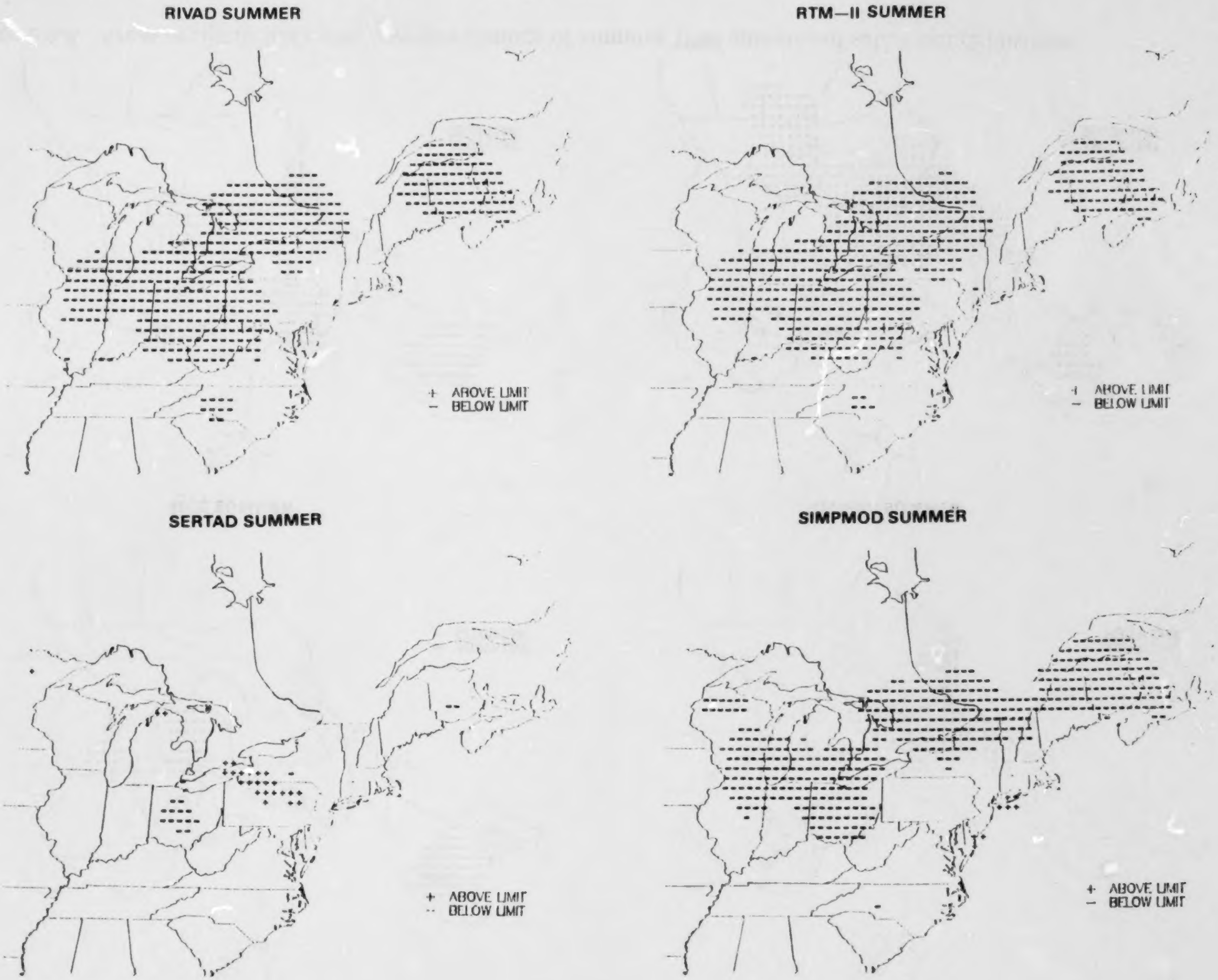


Figure 6.8 Areas of significant under/overpredictions of **summer 1980** amounts of sulfur wet deposition.



RIVAD SUMMER

RTM—II SUMMER

SERTAD SUMMER

SIMPMOD SUMMER

Figure 6.8 Continued

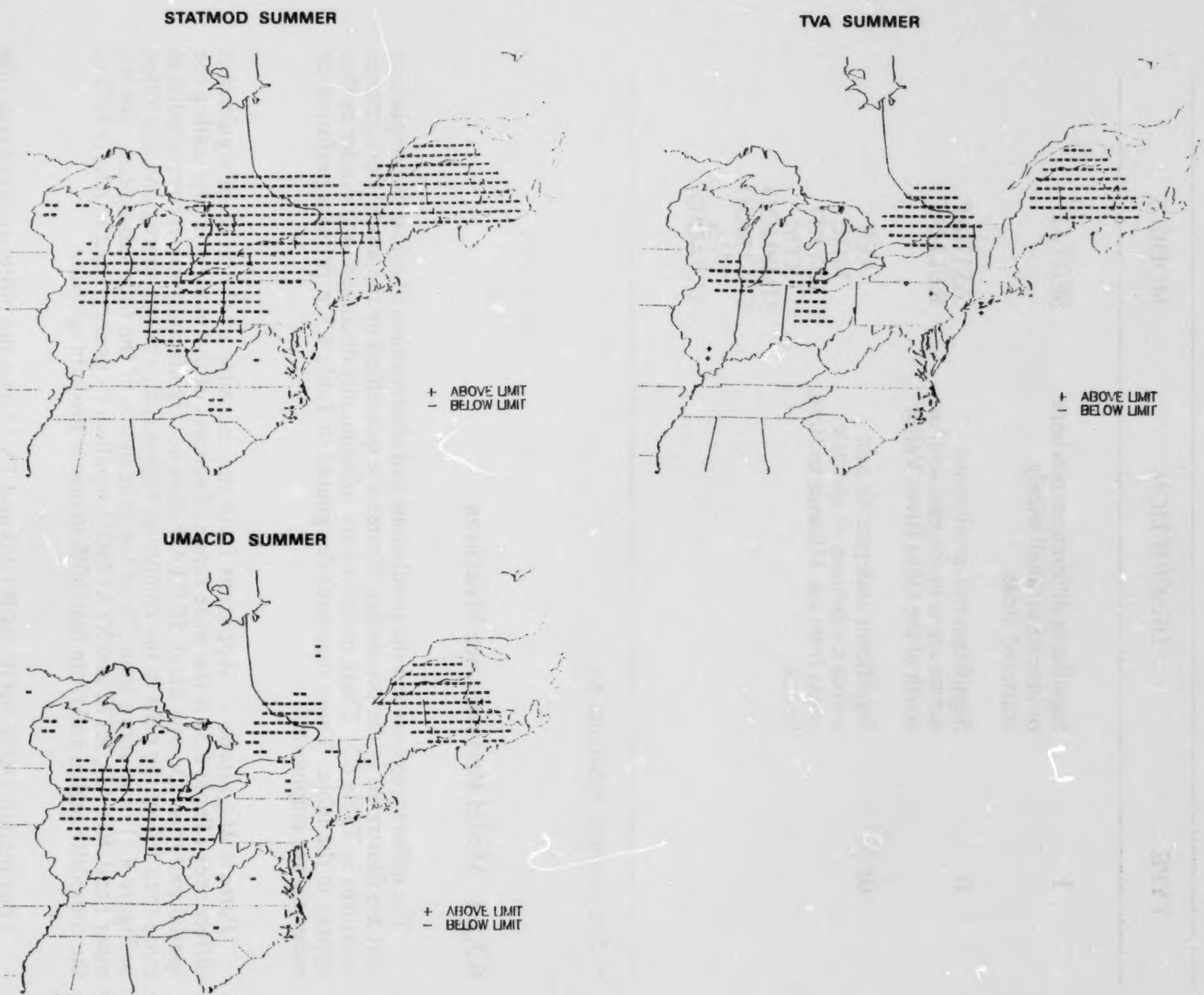


Figure 6.8 Continued

TABLE 6.2. TYPES OF PATTERNS OF SIGNIFICANT DIFFERENCES FOR SUMMER 1980.

TYPE	DESCRIPTION	MODELS
I	Significant differences nowhere or limited to small widely scattered areas	SERTAD*
II	Significant overpredictions across areas to the east and/or south of the Ohio River Valley	ASTRAP RELMAP
III	Significant underpredictions across continuous or sporatic areas from the Midwest to Nova Scotia	AES MOE RIVAD RTM SIMPMOD STATMOD TVA UMACID

* See footnote in Section 6.1.

6.3.1.4 Model Performance Measures

The differences between the predictions and observations of sulfur wet deposition that are illustrated in the preceeding figures are quantified by the model performance measures in Table 6.3. These measures are subsequently discussed in the order as they appear in the table. Those measures designated in Table 4.3 as primary performance measures are emphasized.

Pattern Magnitude - Apparent from the measure of the area of significant differences (Table 6.3) is the wide range of success of the models in replicating the spatial patterns. On one hand, SERTAD does not significantly under/overpredict in more than 5% of the cells in the composite region (i.e., all four subregions illustrated in Figure 4.1) or in more than 5% of the cells in any of the four subregions. On the other hand, RIVAD, RTM and STATMOD significantly underpredict in nearly 40% of the composite region and more than 40% in one or two subregions.

Four models (AES, MOE, SERTAD and TVA) pass the more stringent criterion for the area of significant differences (i.e., significant differences in no more than 20%

TABLE 6.3 SUMMARY OF MODEL PERFORMANCE MEASURES FOR SUMMER 1980.

	AES	AST	MOE	REL	RIV	RTM	SER	SIM	STA	TVA	UMA	OBS
I. Spatial Pattern Replication												
A. Magnitude												
1. Area of Significant Differences (%)*												
a. Northeast	16.9-	1.1-	17.7-	7.5-	26.3-	22.9-	0.4-	29.3-	37.6-	22.3-	27.1-	
b. North Central	4.3-	6.4+	10.0-	5.9+	42.8-	48.0-	3.2+	38.7-	54.8-	16.3-	16.3-	
c. Southwest	14.5-	4.9+	17.1-	32.9+	46.5-	46.9-	4.4-	31.1-	29.8-	10.9-	31.1-	
d. Southeast	3.9+	48.7+	3.3-	90.1+	27.0-	26.3-	3.9+	3.3+	11.2-	2.0+	0.7	
e. All 4 regions	9.4-	10.6+	12.4-	23.7+	37.3-	38.6-	2.9	29.[illegible]	39.2-	14.9-	20.0-	
2. Median (kg S/ha)	2.97	3.43	2.18	4.49	1.70	1.57	3.31	2.13	1.51	2.80	1.66	3.00
(p-value)	0.826	0.510	0.048	0.016	0.000	0.000	0.510	0.048	0.004	0.826	0.048	
3. Dispersion (kg S/ha)	1.63	4.33	1.28	12.07	0.25	0.20	1.11	1.12	0.98	1.83	1.50	2.71
(p-value)	0.055	0.069	0.009	0.000	0.010	0.000	0.003	0.003	0.001	0.105	0.031	
4. Bias-High (kg S/ha)	0.91	-0.84	1.24	-2.01	2.50	2.60	0.40	1.88	2.28	0.87	1.66	
(p-value)	0.01	0.04	0.00	0.03	0.00	0.00	0.13	0.00	0.00	0.02	0.00	
5. Bias-Low (kg S/ha)	0.10	-0.25	0.67	-1.07	0.91	0.92	-0.60	0.61	0.99	0.30	1.04	
(p-value)	0.33	0.17	0.00	0.02	0.00	0.00	0.00	0.01	0.00	0.11	0.00	
6. Mean Squared Error(kg S/ha)^2	1.98	2.71	2.01	11.6	4.11	4.16	1.41	3.41	3.98	2.17	3.23	
B. Position												
1. Centroids (km)	100	63	51	147	33	54	10	127	92	94	42	
2. Major Axis (degrees)**	+2.8	+8.5	+11.7	+10.3	+6.8	+6.2	+14.6	+13.4	+19.6	+12.5	-0.6	40.5
3. Shift (km)	190	220	156	348	220	141	190	453	110	245	156	
C. Shape												
1. Correlation-No Shift	0.56	0.68	0.77	0.38	0.81	0.80	0.75	0.30	0.64	0.54	0.74	
2. Correlation-With Shift	0.88	0.83	0.84	0.71	0.88	0.89	0.88	0.58	0.67	0.74	0.85	
II. Maxima												
A. Location-Kriged (km)	277	351	84	277	84	84	263	259	164	94	277	
B. Magnitude-Kriged (%)**	-9	+1	-34	+85	-60	-59	-1[illegible]	-32	-42	-14	-30	
Magnitude-Site (%)**	-2	+1	-31	+89	-59	-56	-1[illegible]	-27	-39	-7	-27	
III. Temporal Pattern Replication												
A. Seasonal Contribution (%)	32.4	34.8	35.1	42.5	38.1	36.2	36.9	23.6	27.4	34.8	34.1	38.5
(Percent Difference)**	-16	-10	-9	+10	-1	-6	-4	-39	-29	-10	-11	

Underscored measures are primary performance measures.

* "+" and "-" denote the prevalence of overpredictions and underpredictions, respectively.

** Values listed for each model represent the value of the prediction minus the value of the observation. Percent differences are normalized by observed values.

of the composite region and no more than 25% in each subregion). Each receives the maximum score of four for pattern magnitude. With the exception of SERTAD, these models have a tendency to underpredict, especially in the northeast. The significant differences for SERTAD are limited to small widely scattered areas.

MOE is only one of two models, SERTAD being the other, which superimposes a background contribution to the predictions of sulfur wet deposition (Tables 2.1 and 2.2). Although each model considers background contributions, the definitions of the background levels for the two models are quite different. The MOE background contribution attempts to account for the effects of the sulfur emissions from biogenic sources inside the model domain, which were not included in the anthropogenic emissions inventory, as well as from sources outside the model domain. A total annual background contribution of 0.7 kg S/ha is added to the predicted amounts across each cell in the model domain. This amount is about 30% of the observed annual amounts at the remote ISDME sites, where background levels are relatively significant, and less than 10% of those at sites in the high deposition region.

On the other hand, the SERTAD background contribution is comprised of two components; one is similar to the MOE background. The other component assumes that a minimum of 80% of the sulfur emissions from stacks exceeding 250 m is wet deposited across the model domain. The amount deposited across a cell is proportional to only the precipitation amount for that cell; source-receptor distances are irrelevant to the amount deposited. Hence, the spatial pattern of SERTAD predictions of sulfur wet deposition, which for a statistical model would normally be rather smooth, would tend to resemble the precipitation pattern. Unlike the background levels for MOE, those for SERTAD were much higher. For example, based on typical precipitation rates across Ohio during the summer of 1980, the SERTAD background contributions for summer are 2.4 kg S/ha, or about 40% of the observed amount.

With the exception of ASTRAP, all of the other models under/overpredict in more than 20% of the composite region and in more than 25% of two or more subregions. These models, therefore, receive a score of zero for pattern magnitude. ASTRAP satisfies the less stringent criterion and, therefore, receives a score of two. The relatively high ASTRAP variance probably is not a factor in its score, since it is unlikely that the model would have otherwise significantly overpredicted in twice as many cells in the composite region (i.e., 20% rather than 10%).

For this season, the assessment of model performance based only on the measure of the area of significant differences generally concurs with that based on only the descriptive statistics. Three of the four models with the maximum score for the pattern magnitude, MOE excluded, have medians and dispersions closest to those of the observations and have the lowest biases and mean squared errors. Furthermore, those models failing the criteria for the area of significant differences also perform least favorably according to the descriptive statistics alone.

However, one must not rely on descriptive statistics alone. There are two good examples for this season. First, the spatial analysis of the MOE predictions indicates

that the model performs rather well in replicating the pattern, but the descriptive statistics give the appearance that the model performs relatively poorly (i.e., the median and dispersion are lower and the biases are higher than those for AES, SERTAD and TVA). Secondly, the descriptive statistics for ASTRAP indicate that this model performs very well, however, the spatial analysis indicates that this model significantly overpredicts across nearly half of the southeast subregion.

Pattern Position - **Four models (ASTRAP, RIVAD, RTM and UMACID) pass both criteria for pattern position. Their centroids (Figure 6.9) are within 65 km of the observed centroids and the orientation of their major axes differs from that of the observations by less than 10 degrees (Table 6.3). Figure 6.9 indicates that the centroids of all models but MOE and SERTAD are located to the south and east of the observed centroid. This shift is indicative of the tendency of the models to predict the greatest values to the south and east of the greatest observations. The MOE centroid is the only one to the north of the observed centroid, while the SERTAD centroid is virtually colocated. The difference in the location of the MOE centroid relative to those predicted by the other models is consistent for all seasons and could reflect the method in which this model simulates transport. This is the only model which uses geostrophic wind velocities calculated from surface pressure fields to simulate transport.**

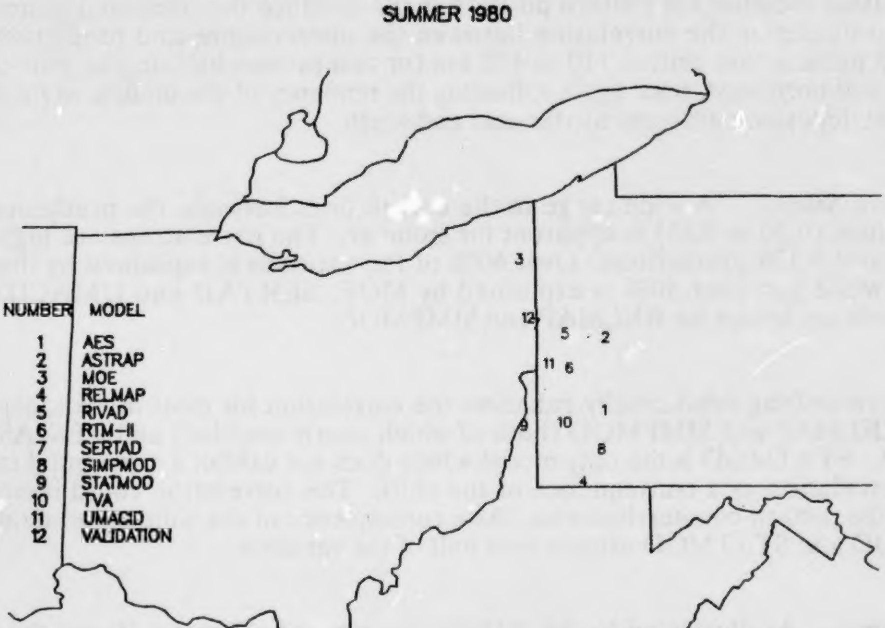


Figure 6.9 **Summer 1980** locations of the centroids.

Four models (RELMAP, SIMPMOD, STATMOD and TVA) pass neither criterion for pattern position. Their centroids are at least 90 km from the observed centroids and the orientation of their major axes differs from that of the observations by as much as 20 degrees.

The four statistical models (SERTAD, SIMPMOD, STATMOD and TVA) exhibit the largest differences in the orientation of the major axis. Unlike that for the other models, the orientations for the statistical models tends to be more west-to-east than the observed southwest-to-northeast orientation. The greater differences for the statistical models could be an artifact of the methods in which the transport winds are determined. These models compute wind roses to represent the long-term tendencies of the boundary layer wind flow throughout the simulation period.

A potentially serious problem with this approach is illustrated. During the summer, the long-term average flow across much of the northeastern United States tends to be from the west. However, prior to and during precipitation events, if the flow in the boundary layer tends to be from the southwest, the statistical models would tend to transport and scavenge sulfur to the east of the major source regions, while the deterministic models, which use actual wind measurements, would tend to transport and scavenge sulfur to the northeast of the sources.

The final measure for pattern position is the distance the predicted patterns are shifted to maximize the correlation between the observations and predictions. All predicted patterns are shifted 110 to 453 km (or two to nine half-degree grid cells) to the west and northwest, once again reflecting the tendency of the models to predict the higher wet deposition amounts too far east and south.

Pattern Shape - A wide range in the correlations between the predictions and observations (0.30 to 0.81) is apparent for summer. The correlations are highest for RIVAD and RTM predictions. Over 60% of the variance is explained by these two models, while just over 50% is explained by MOE, SERTAD and UMACID. The correlations are lowest for RELMAP and SIMPMOD.

Pattern shifting substantially enhances the correlation for most models, especially that of RELMAP and SIMPMOD (both of which nearly doubled) and AES, ASTRAP and TVA. STATMOD is the only model which does not exhibit a substantial increase in the correlation as a consequence of the shift. The correlation could increase by rotating the pattern counterclockwise. As a consequence of the shifting, all models but SIMPMOD and STATMOD explain over half of the variance.

Maxima - As illustrated by the ISDME site values in Figure 6.10 and the spatial patterns depicted in Figure 6.7, the models generally predict the maximum in or near West Virginia, while the maximum observation is located in central Ohio. Only TVA passes the criterion for both the location and magnitude of the maximum value. Its maximum prediction was within 85 km of the location and within 15% of the magnitude of the observed maximum. Conversely, RELMAP, SIMPMOD and STATMOD pass neither criterion.

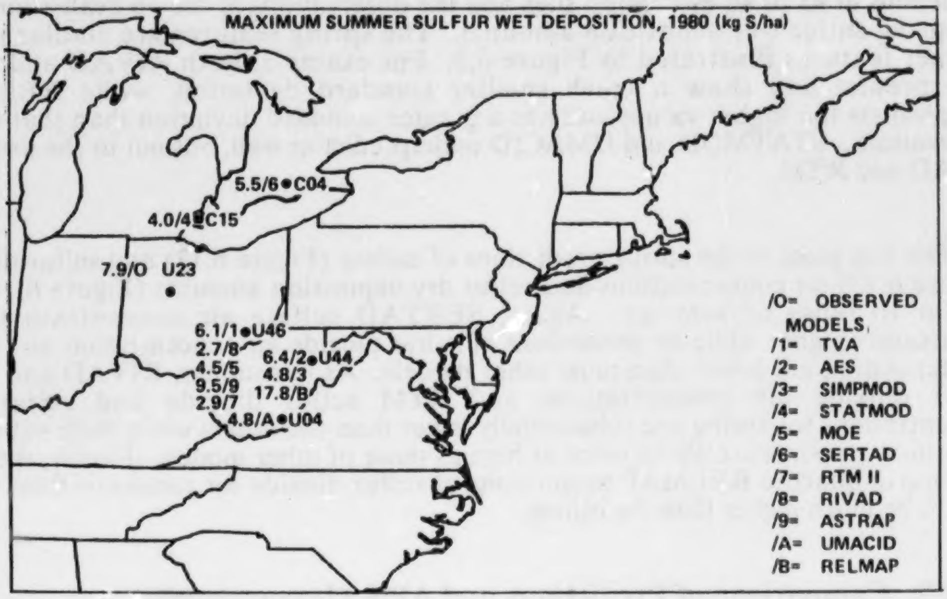


Figure 6.10 **Summer 1980** locations and magnitudes (kg S/ha) of the maximum amount of sulfur wet deposition at the ISDME sites.

Three models (MOE, RIVAD and RTM) pass the criterion for the location of the maximum, but each underpredicts the magnitude by 30 to 60%. These underpredictions are consistent with the general tendencies of these models to underpredict everywhere.

Four other models (AES, ASTRAP, SERTAD and UMACID) pass the criterion for the magnitude, but fail the criterion for the location. These models predict the location of the maximum value at least 250 km from the nearest probable location.

Temporal Pattern - Only AES, SIMPMOD and STATMOD fail the criterion for the summer contribution to the annual predicted total. As illustrated by Figure 6.1, the percent contribution for each of these models is 15 to 40% lower than that observed, partly as a result of their high winter contributions.

6.3.2 Spring 1980 Results

6.3.2.1 Summary of Distributional Statistics

Illustrated by Figure 6.11 are the statistical features of the distributions of the predictions at 41 to 46 evaluation sites and the observations at the 46 evaluation sites for spring sulfur wet deposition amounts. The spring features are similar to the summer features illustrated by Figure 6.2. For example, both RIVAD and RTM underpredict and show a much smaller standard deviation, while RELMAP overpredicts the higher values and has a greater standard deviation than that of the observations. STATMOD and UMACID underpredict as well, but not to the extent of RIVAD and RTM.

The box plots of the spring predictions of sulfate (Figure 6.12) and sulfur dioxide (Figure 6.13) air concentrations and sulfur dry deposition amounts (Figure 6.14) are similar to those of summer. Again, SERTAD sulfate air concentrations are significantly higher while its predictions of sulfur dioxide air concentration and sulfur dry deposition are lower than most other models. As in summer, RIVAD and RTM sulfur dioxide air concentrations and RTM sulfur dioxide and sulfate air concentrations for spring are substantially lower than the others while their sulfur dry deposition amounts are about twice as high as those of other models. Finally, the third and fourth quartile RELMAP predictions of sulfur dioxide air concentrations again tend to be much higher than the others.

6.3.2.2 Comparison of Predictions and APN Measurements

Illustrated in Figure 6.15 are the spring averages of the predicted and observed air concentrations of sulfur dioxide and sulfate at the APN sites. Similar to the summer comparisons, the four statistical models predict the highest sulfate concentrations and overpredict the sulfate concentrations for spring. SERTAD sulfate concentrations are as much as six times too high, while those of the other models are generally within 50% of the observations.

Also similar to the summer comparisons, most models underpredict the sulfate and overpredict the sulfur dioxide concentrations at C05, the Nova Scotia site. The reason for this is not apparent. For the sulfur dioxide concentrations, RIVAD overpredicts at each site and RELMAP overpredicts at C04 and C05, the sites which are more often downwind of the major source regions. This is also the case for summer.

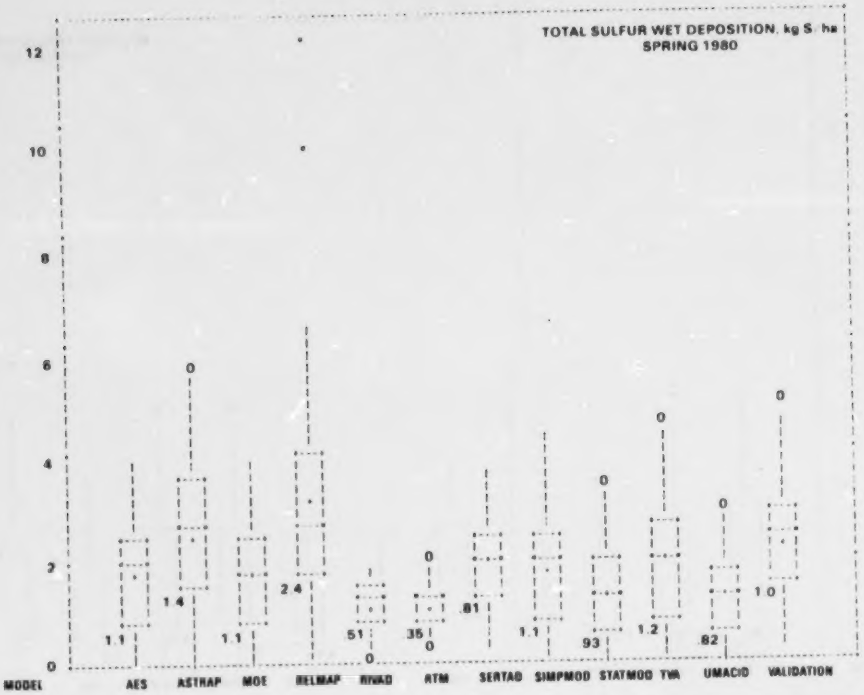


Figure 6.11 **Spring 1980** box plots of the predictions and observations of sulfur wet deposition (kg S/ha).

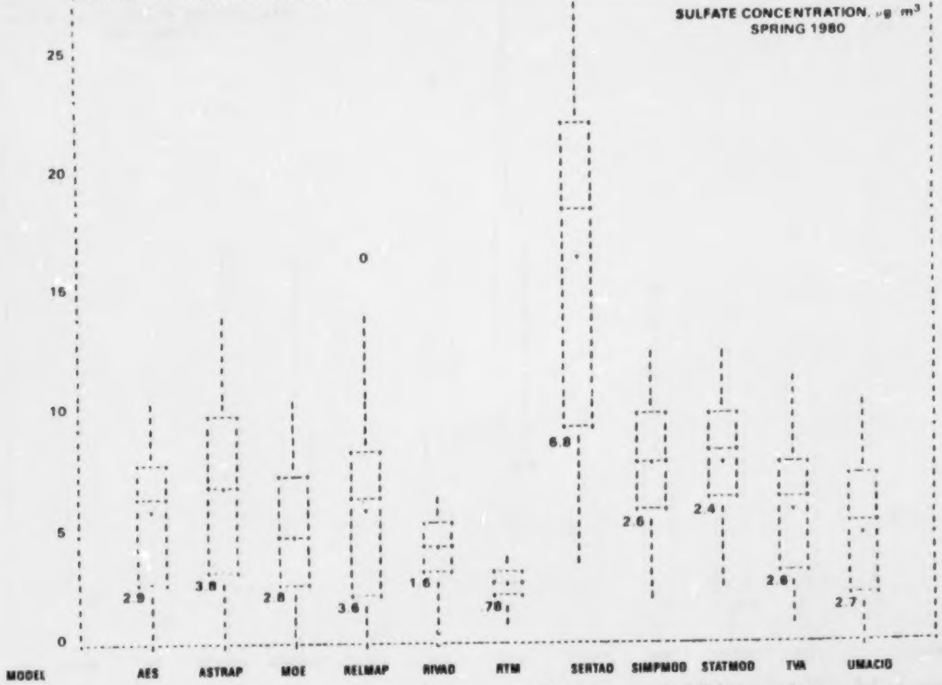


Figure 6.12 **Spring 1980** box plots of the predictions of sulfate air concentrations (µg/m³).

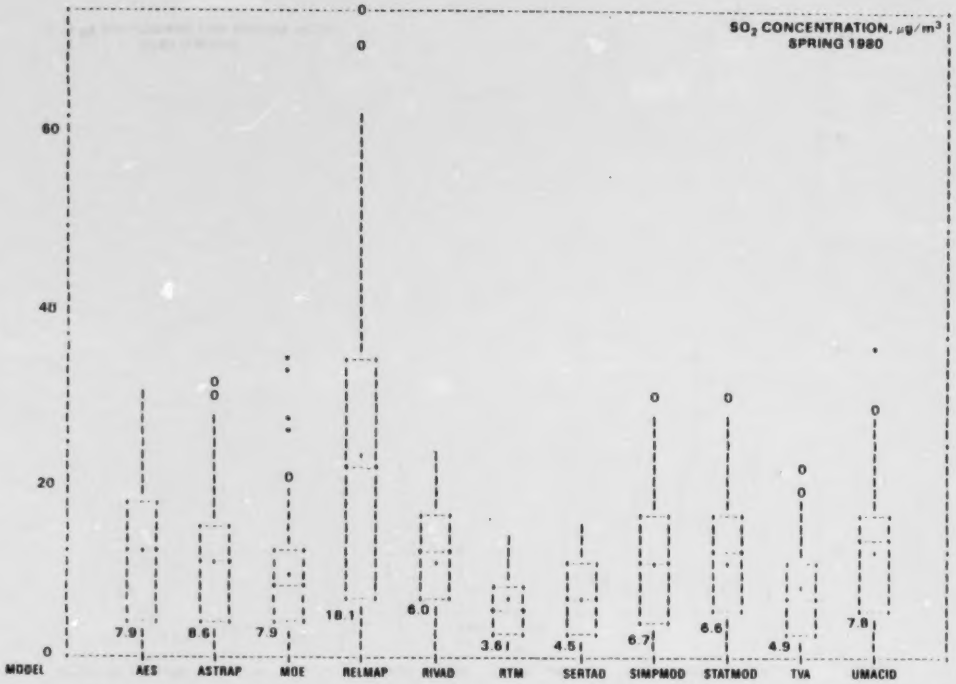


Figure 6.13 **Spring 1980** box plots of the predictions of sulfur dioxide air concentrations (µg/m^3).

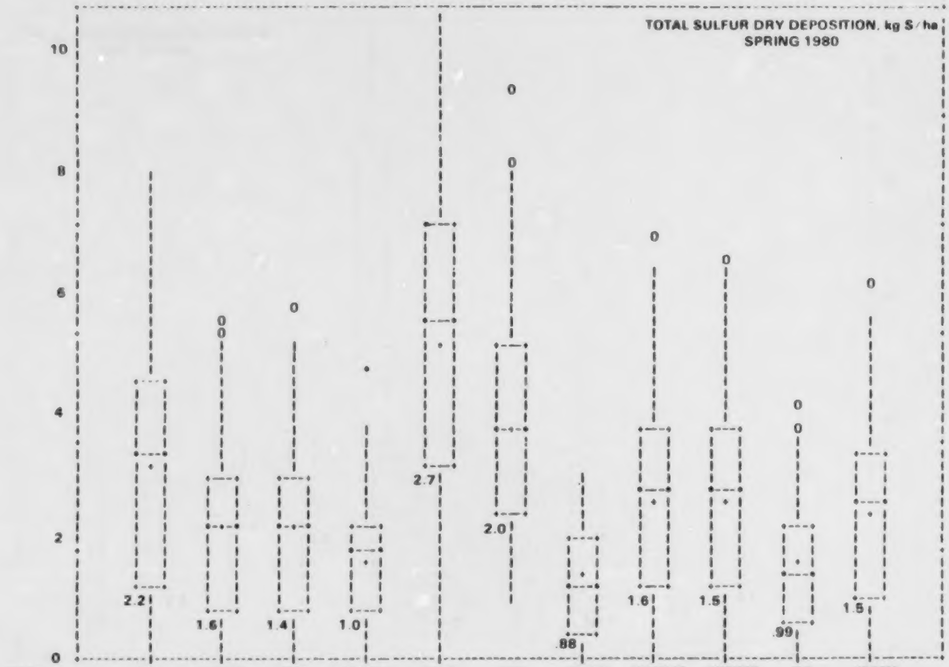


Figure 6.14 **Spring 1980** box plots of the predictions of sulfur dry deposition (kg S/ha).

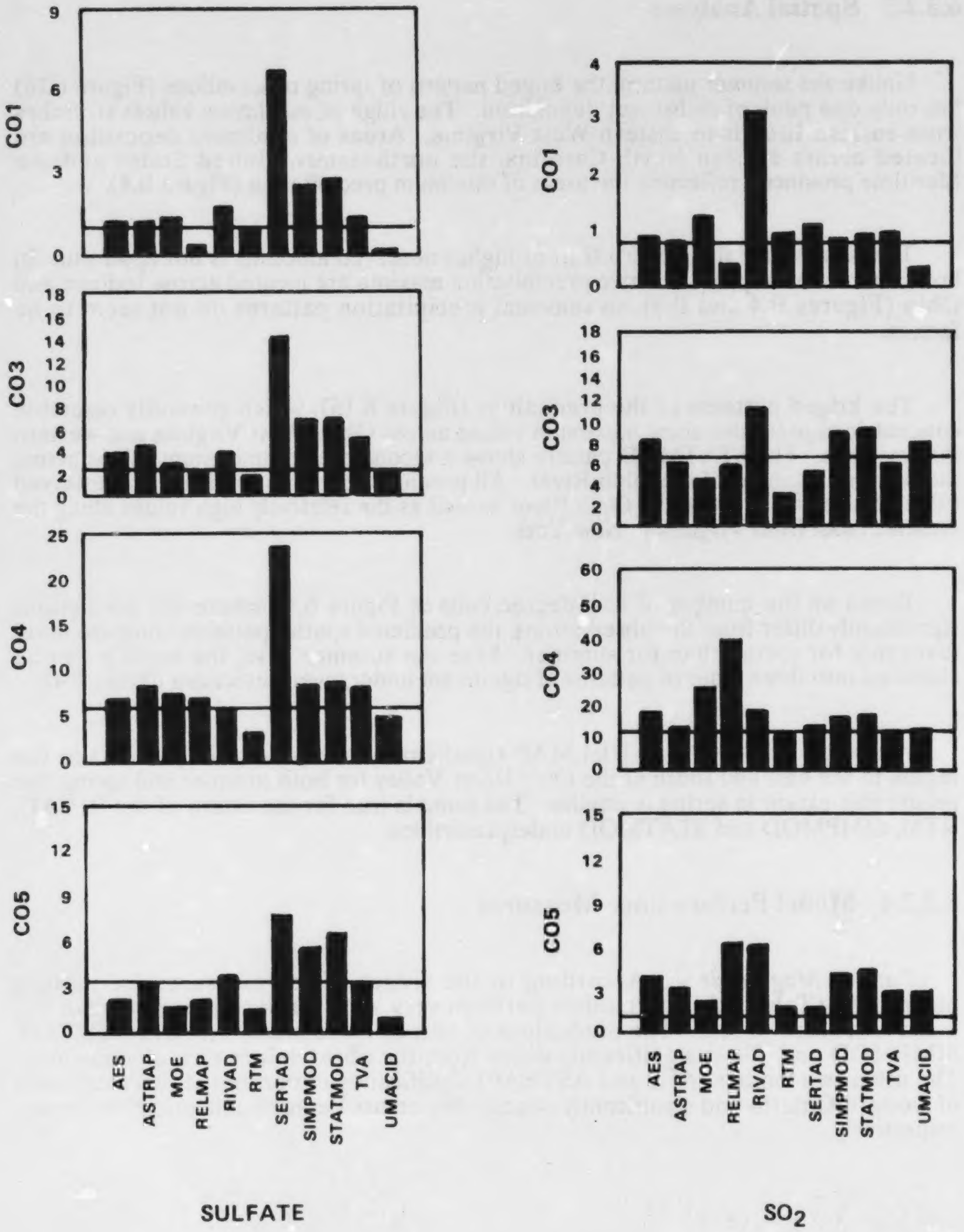


Figure 6.15 Comparison of the **spring 1980** observations and predictions of the air concentrations ($\mu g/m^3$) of sulfur dioxide and sulfate at the APN sites.

6.3.2.3 Spatial Analyses

Unlike the summer pattern, the kriged pattern of spring observations (Figure 6.16) has only one peak of sulfur wet deposition. The ridge of maximum values stretches from eastern Illinois to eastern West Virginia. Areas of minimum deposition are located across eastern North Carolina, the northeastern United States and the Maritime provinces, reflecting the areas of minimum precipitation (Figure B.4).

The reasons for the eastern shift of higher observed amounts is not apparent. In both summer and spring, relative precipitation maxima are located across Indiana and Ohio (Figures B.4 and B.5), so seasonal precipitation patterns do not seem to be factors.

The kriged patterns of the predictions (Figure 6.16), which generally resemble concentric elipses, also show maximum values across Ohio, West Virginia and western Pennsylvania. The STATMOD pattern shows a second area of maximum values across the western segment of the Ohio River. All predicted patterns replicate the observed ridge of high values along the Ohio River as well as the relatively high values along the Atlantic coast from Virginia to New York.

Based on the number of half-degree cells in Figure 6.17 where the predictions significantly differ from the observations, the predicted spatial patterns compare more favorably for spring than for summer. Like the summer case, the models can be classified into three types of patterns of significant under/overpredictions (Table 6.4).

Although ASTRAP and RELMAP significantly overpredict in portions of the region to the east and south of the Ohio River Valley for both summer and spring, the geographic extent in spring is smaller. The same is true for the extent of the RIVAD, RTM, SIMPMOD and STATMOD underpredictions.

6.3.2.4 Model Performance Measures

Pattern Magnitude - According to the values of the measures for pattern magnitude (Table 6.5), six models perform very well for spring and receive the maximum score of four. The predictions of four of these models (MOE, SERTAD, SIMPMOD and TVA) significantly differ from the observations virtually nowhere. The other two models (AES and ASTRAP) significantly underpredict in a small area of western Ontario and significantly overpredict across Pennsylvania and New Jersey, respectively.

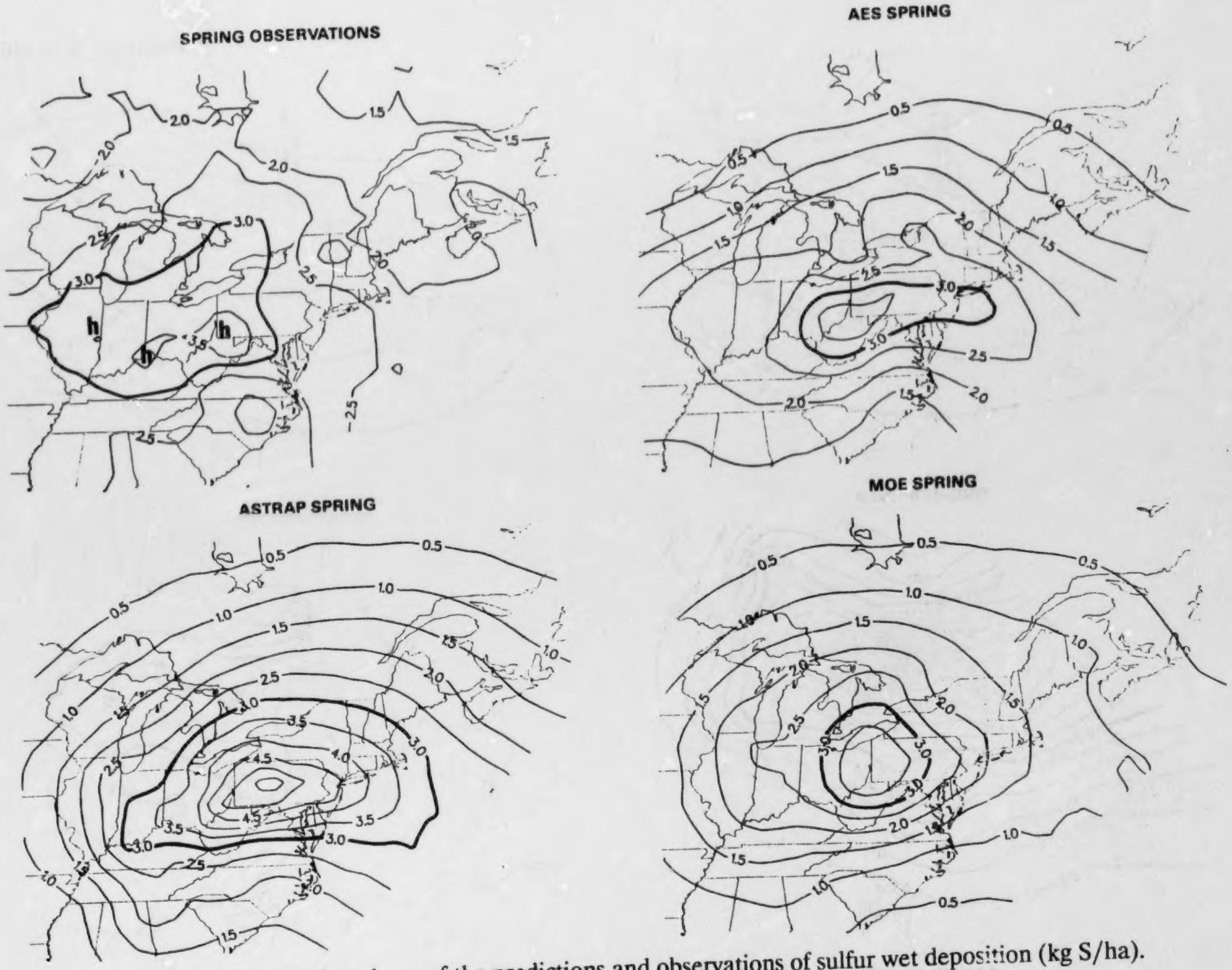
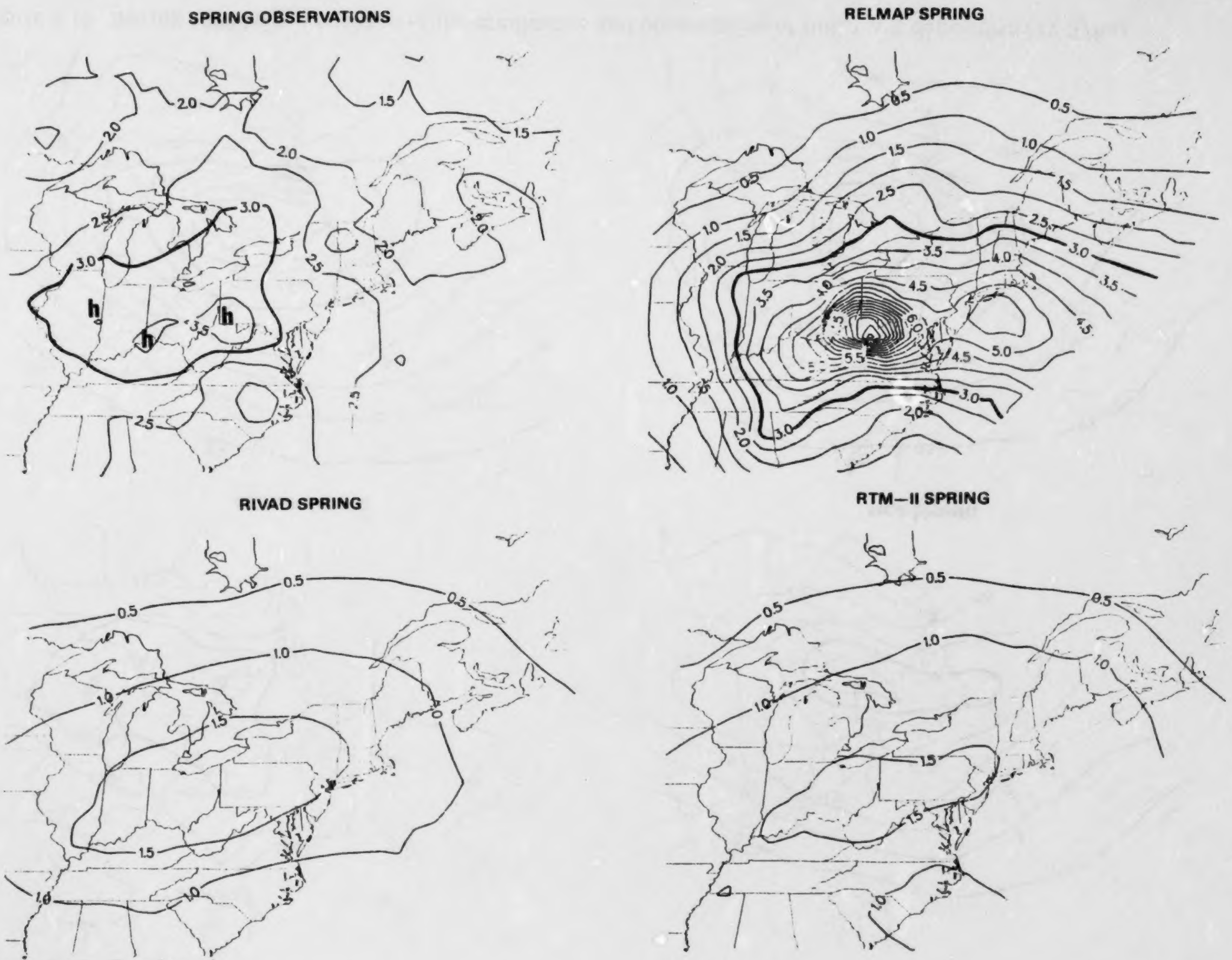


Figure 6.16 **Spring 1980** kriged analyses of the predictions and observations of sulfur wet deposition (kg S/ha).



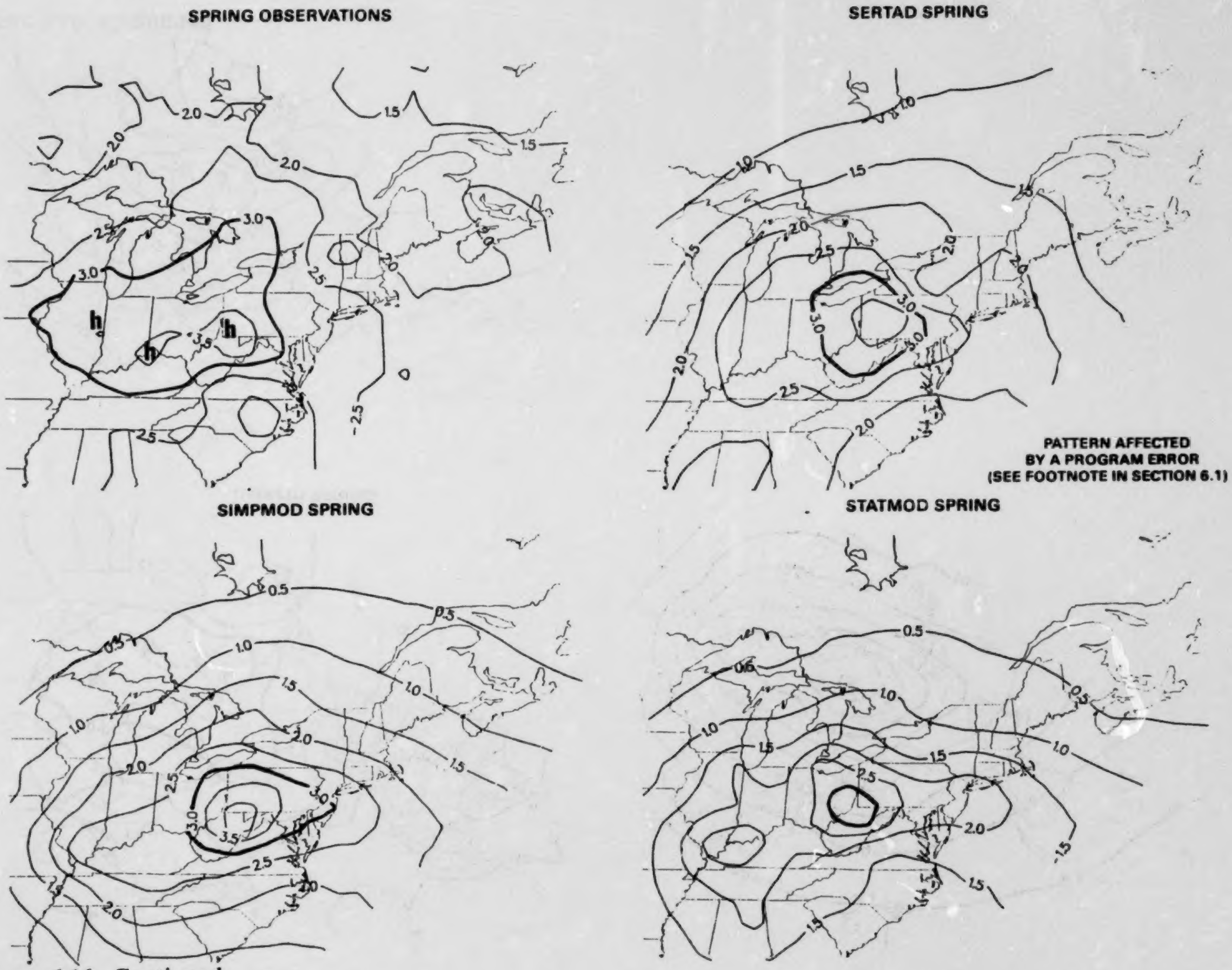
SPRING OBSERVATIONS

RELMAP SPRING

RIVAD SPRING

RTM—II SPRING

Figure 6.16 Continued



SPRING OBSERVATIONS

SERTAD SPRING

PATTERN AFFECTED
BY A PROGRAM ERROR
(SEE FOOTNOTE IN SECTION 6.1)

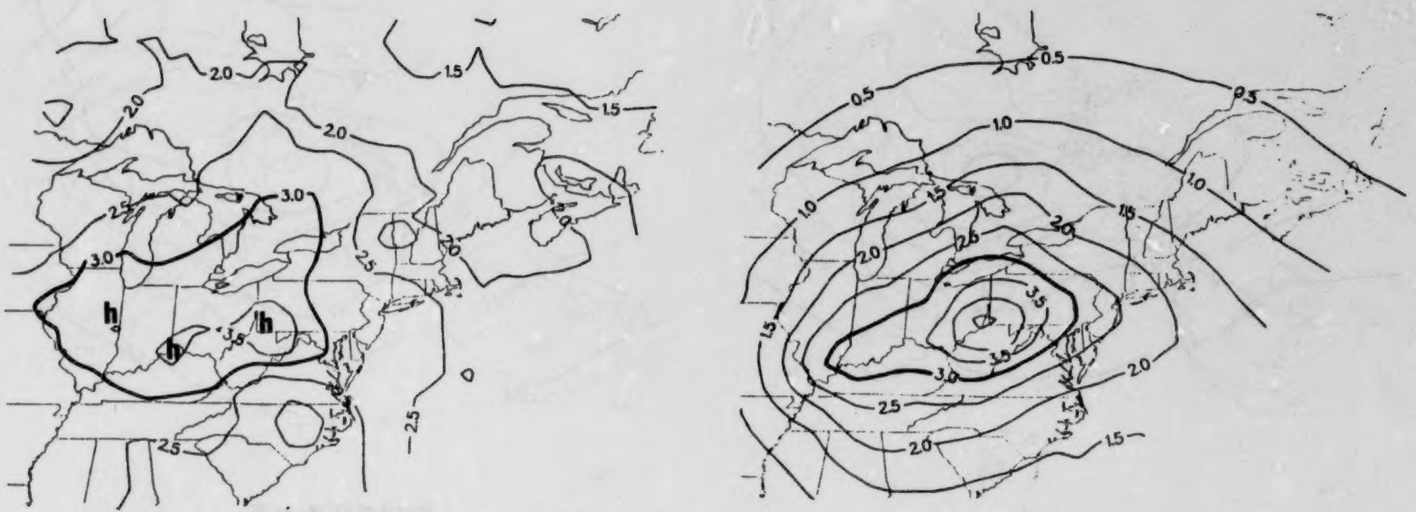
SIMPMOD SPRING

STATMOD SPRING

Figure 6.16 Continued

SPRING OBSERVATIONS

TVA SPRING



UMACID SPRING

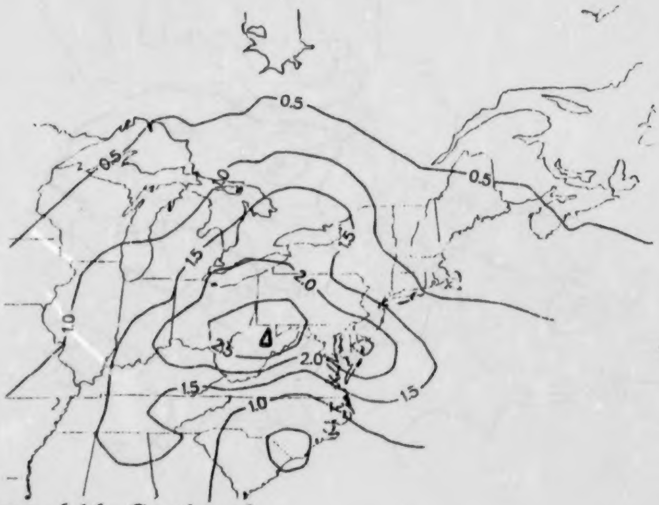


Figure 6.16 Continued

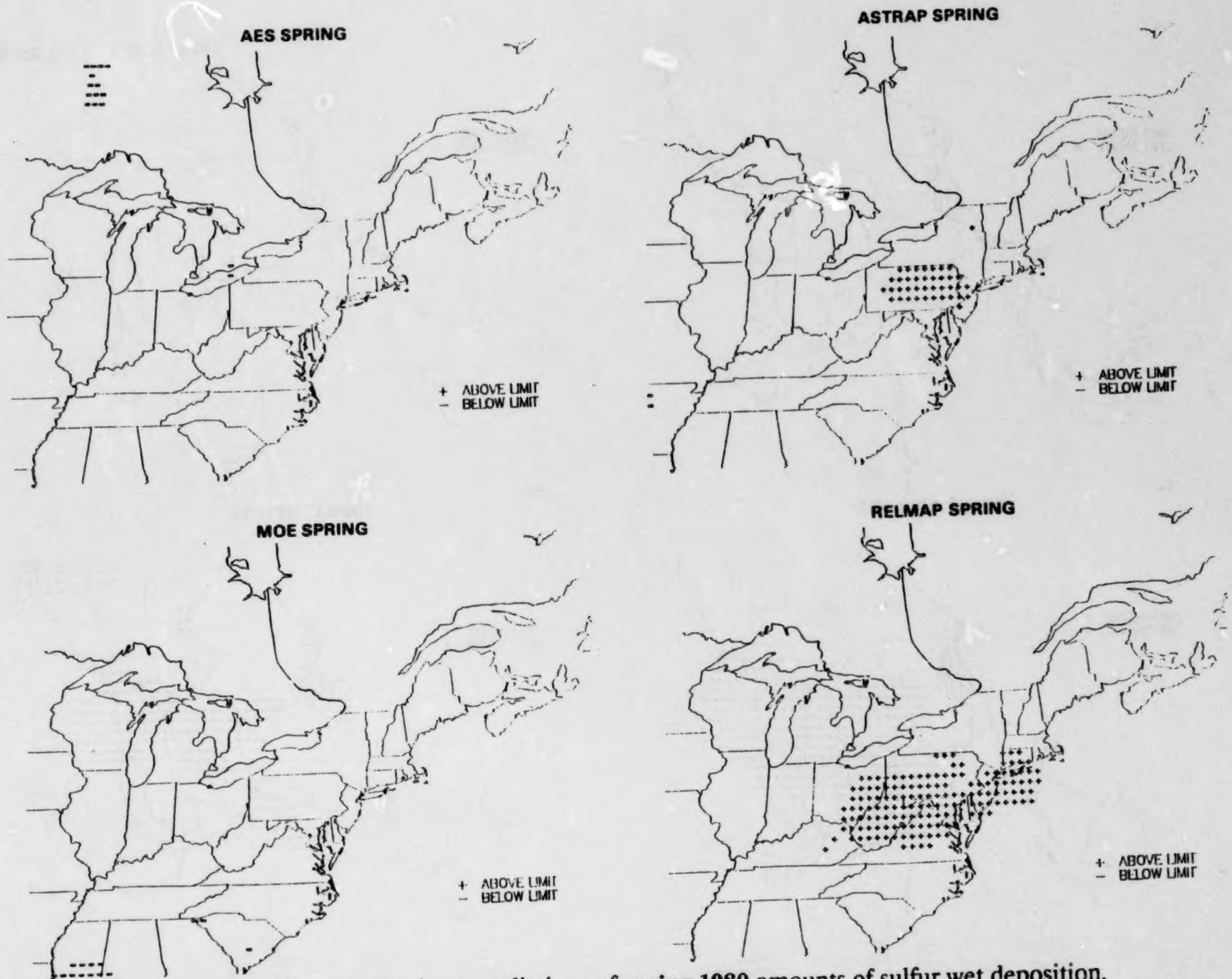


Figure 6.17 Areas of significant under/overpredictions of **spring 1980** amounts of sulfur wet deposition.

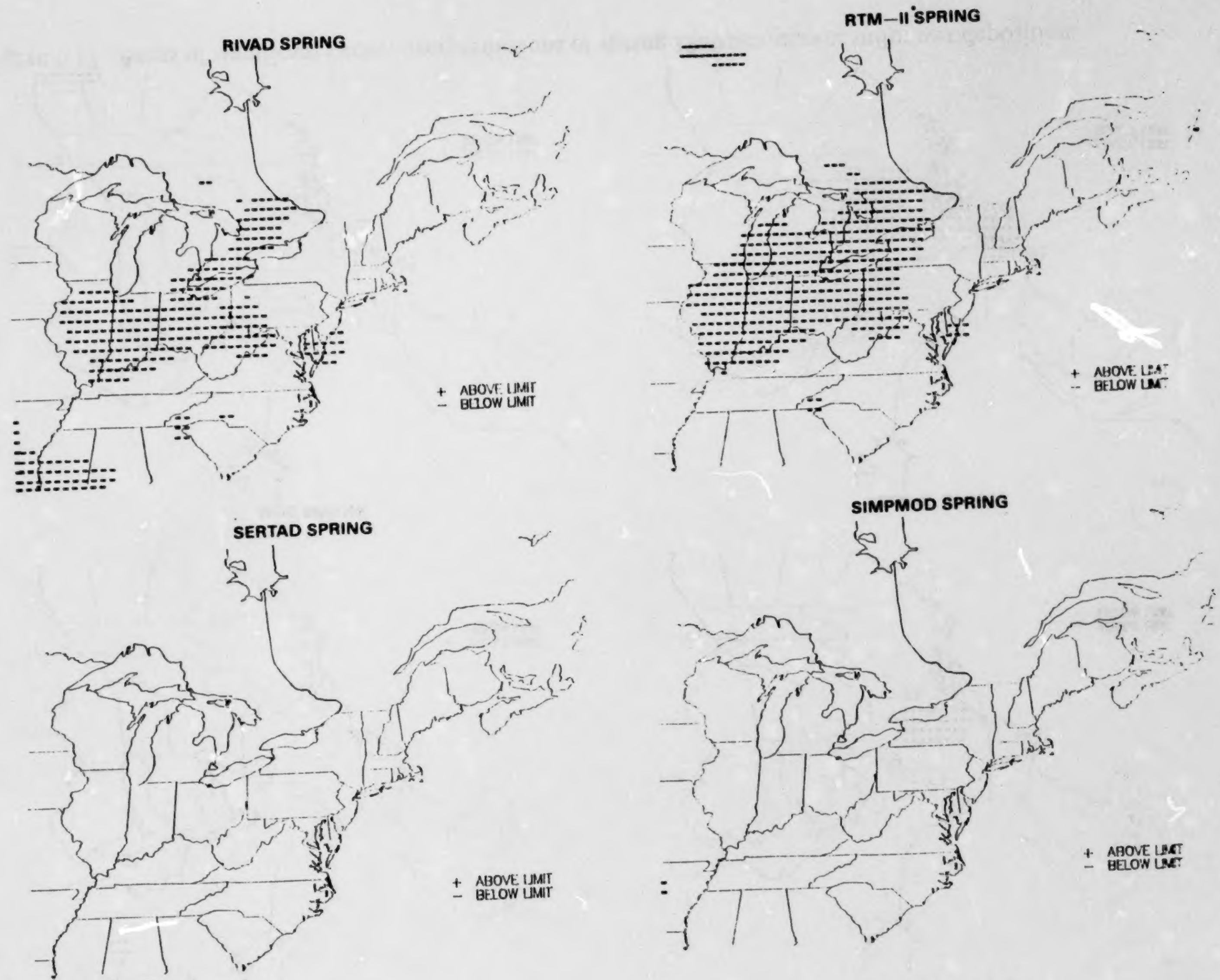


Figure 6.17 Continued

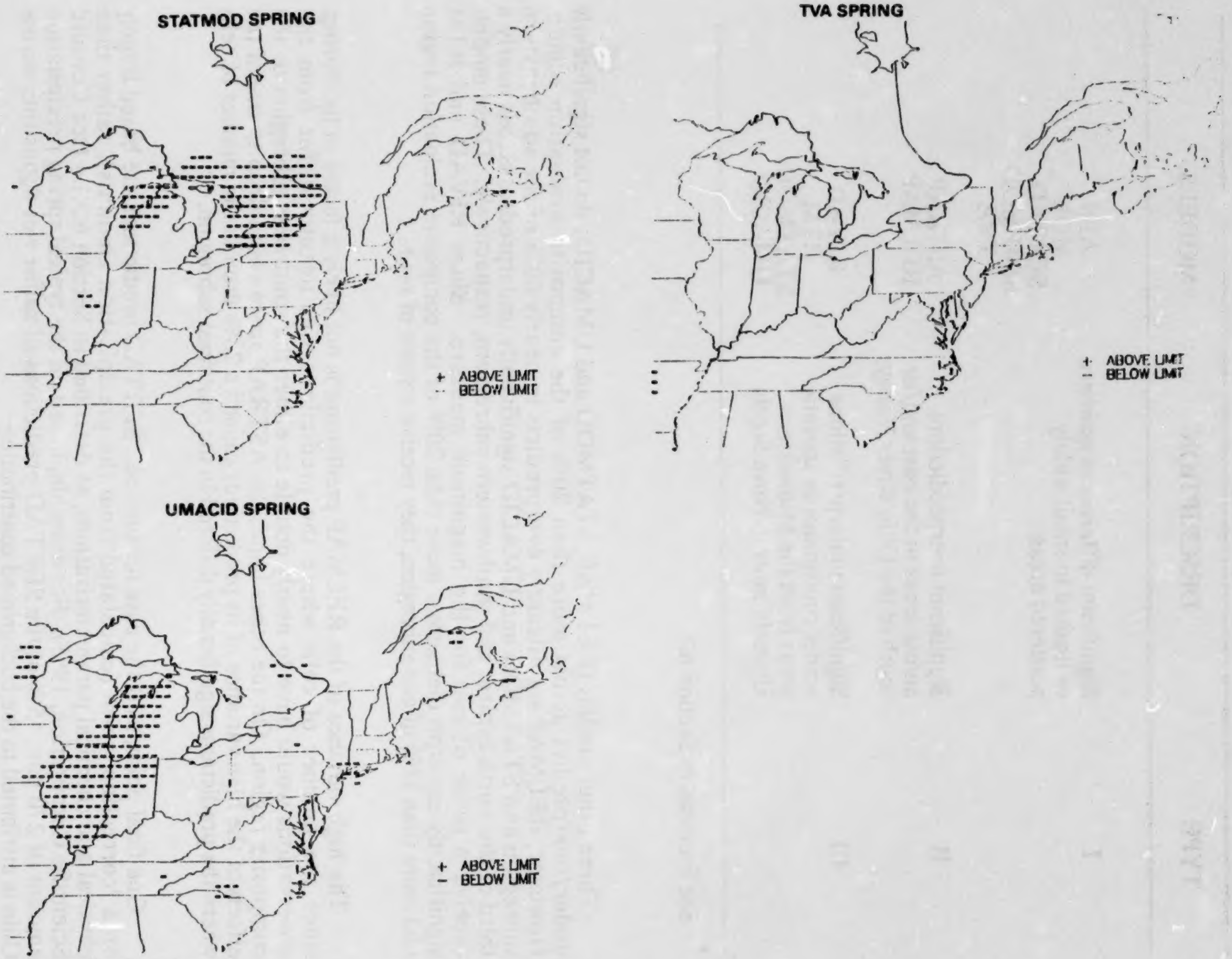


Figure 6.17 Continued

TABLE 6.4 TYPES OF PATTERNS OF SIGNIFICANT DIFFERENCES FOR SPRING 1980.

TYPE	DESCRIPTION	MODELS
I	Significant differences nowhere or limited to small widely scattered areas	AES MOE SERTAD* SIMPMOD TVA
II	Significant overpredictions across areas to the east and/or south of the Ohio River Valley	ASTRAP RELMAP
III	Significant underpredictions across continuous or sporatic areas from the Midwestern United States to Nova Scotia	RIVAD RTM STATMOD UMACID

* See footnote in Section 6.1.

Three other models (RELMAP, STATMOD and UMACID) do not significantly under/overpredict across more than 20% of the composite evaluation region. However, RELMAP significantly overpredicts in nearly 60% of the southeastern subregion and STATMOD and UMACID significantly underpredict across nearly a third of the north central and southwestern subregions, respectively. These models receive a score of two for the magnitude measure. Since RIVAD and RTM significantly underpredict across more than 20% of the composite evaluation region and more than 55% of one subregion, they receive a score of zero.

The high variance of the RELMAP predictions is not likely a factor in its scoring, since the number of cells where the predictions significantly differ from the observations would have to nearly double in either the southwest subregion or the composite region. On the other hand, the ASTRAP score could drop to one if the effects of the high variance of its predictions caused a 25% drop in the number of cells where the predictions significantly differed in the southeast subregion.

One final point, as is the case for summer, SERTAD predictions are based largely on a "correction" term (calculated from the precipitation amounts), rather than physical and chemical parameterizations, as described in Section 6.3.1.4 (see Concord Scientific Corporation, 1985). For example, based on the typical spring precipitation amount of 270 mm, 45% of the SERTAD predictions of sulfur wet deposition across Ohio is attributed to the background contribution.

TABLE 6.5 SUMMARY OF MODEL PERFORMANCE MEASURES FOR SPRING 1980.

	AES	AST	MOE	REL	RIV	RTM	SER	SIM	STA	TVA	UMA	OBS
I. Spatial Pattern Replication												
A. Magnitude												
1. Area of Significant Differences (%)*												
a. Northeast	0.0	0.0	0.0	0.0	0.0	0.4-	0.0	0.0	1.9-	0.0	2.3-	
b. North Central	0.5-	1.8+	0.0	1.4+	16.1-	39.1-	0.0	0.0	33.0-	0.0	12.9-	
c. Southwest	0.0	0.0	0.0	14.0+	57.5-	64.0-	0.0	0.0	1.8-	0.0	32.0-	
d. Southeast	0.0	20.4+	0.0	59.9+	25.7-	27.6-	0.0	0.0	0.0	0.0	3.3-	
e. All 4 regions	0.2-	3.6+	0.0	11.9+	22.2-	33.3-	0.0	0.0	14.2-	0.0	13.0-	
2. Median (kg S/ha)	2.20	3.00	2.13	3.07	1.40	1.33	2.33	2.15	1.60	2.25	1.34	2.83
(p-value)	0.124	0.510	0.048	0.272	0.000	0.000	0.079	0.124	0.000	0.124	0.000	
3. Dispersion (kg S/ha)	0.80	1.53	0.90	5.5	0.15	0.10	0.53	0.90	0.76	1.11	0.67	0.92
(p-value)	0.316	0.055	0.456	0.000	0.000	0.000	0.038	0.461	0.268	0.285	0.154	
4. Bias-High (kg S/ha)	0.76	-0.37	0.55	-1.39	1.77	1.93	0.64	0.72	1.33	0.42	1.44	
(p-value)	0.00	0.10	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.03	0.00	
5. Bias-Low (kg S/ha)	0.49	-0.09	0.68	-0.51	1.03	1.85	0.21	0.45	0.91	0.46	1.22	
(p-value)	0.00	0.32	0.00	0.05	0.00	0.00	0.05	0.00	0.00	0.00	0.00	
6. Mean Squared Error(kg S/ha)^2	0.93	1.23	1.00	4.32	2.34	2.52	0.65	0.89	1.64	0.77	2.13	
B. Position												
1. Centroids (km)	51	68	66	82	30	25	16	72	84	46	30	
2. Major Axis (degrees)**	-2.6	-3.1	-0.3	-0.5	-2.2	-2.5	-1.3	-1.6	+7.2	-0.3	+0.2	51.9
3. Shift (km)	347	440	110	156	220	550	110	156	110	156	116	
C. Shape												
1. Correlation-No Shift	0.46	0.52	0.77	0.52	0.69	0.57	0.74	0.51	0.65	0.66	0.80	
2. Correlation-With Shift	0.70	0.73	0.79	0.63	0.76	0.74	0.76	0.71	0.67	0.78	0.83	
II. Maxima												
A. Location-Kriged (km)	0	0	0	0	0	0	0	0	0	0	0	
B. Magnitude-Kriged (%)**	0	0	0	+55	-35	-33	0	0	0	0	0	
Magnitude-Site (%)**	-7	0	-9	+98	-55	-53	-11	0	-17	0	-27	
III. Temporal Pattern Replication												
A. Seasonal Contribution (%)	23.2	26.7	31.2	32.1	30.4	27.7	24.7	22.1	23.7	28.6	22.7	29.3
(Percent Difference)**	-21	-9	+6	+10	+4	-5	-16	-25	-19	-2	-23	

Underscored measures are primary performance measures.

* "+" and "-" denote the prevalence of overpredictions and underpredictions, respectively.

** Values listed for each model represent the value of the prediction minus the value of the observation. Percent differences are normalized by observed values.

Pattern Position - With few exceptions, the models perform equally well in predicting pattern position. The distances between observed and predicted centroids, illustrated by Figure 6.18, are less than 85 km. All models but RELMAP and STATMOD pass the criterion for the centroid.

The orientation of the major axis for the observations (52 degrees) is more west to east in spring than for any other season when the orientation tends to be south-southwest to north-north-east. The orientation of the major axes of the predictions deviates from that of the observations by no more than 8 degrees. Therefore, all models pass the criterion for the orientation.

For all but three models (AES, ASTRAP and RTM), the distance of the pattern shift to maximize the correlation is less than 160 km, slightly larger than one model grid cell. The patterns of the three models are shifted from 350 to 550 km to maximize the correlation.

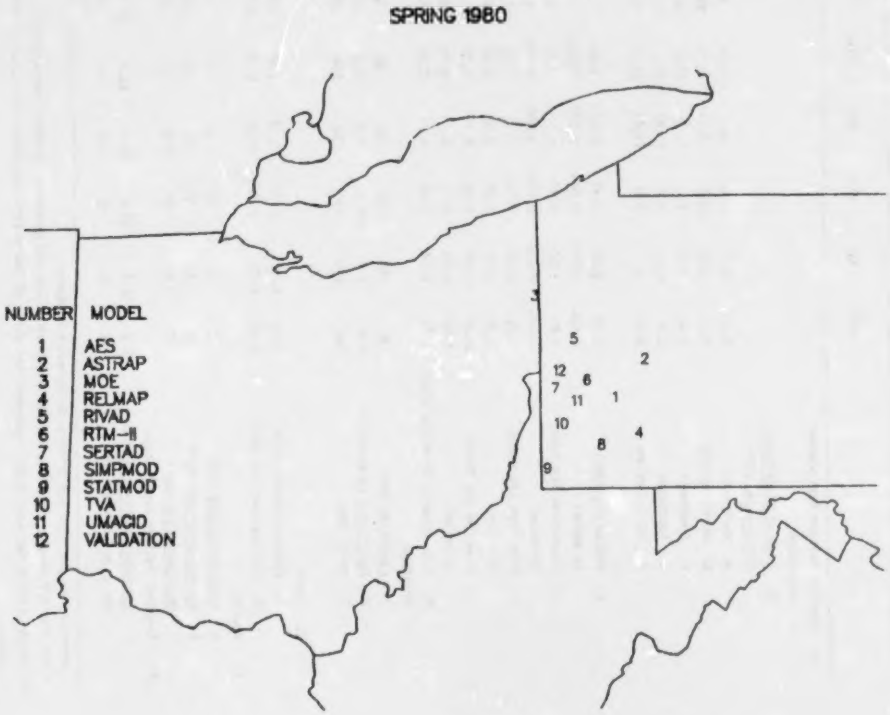


Figure 6.18 **Spring 1980** locations of the centroids.

Pattern Shape - As in the summer, MOE, SERTAD and UMACID have spring correlations in excess of 0.7, indicating a consistency of relatively good model performance in predicting the pattern shape and position for those two seasons. On the other hand, RIVAD and RTM correlations for the spring are considerably lower than the 0.8 correlations for the summer. RELMAP and SIMPMOD correlations are again low (i.e., less than 0.6), as are those for AES, ASTRAP and RTM.

Pattern shifting increases all correlations to nearly 0.7. The magnitude of the increases appear to be correlated to the distance the patterns are shifted to obtain the maximum correlation (i.e., the correlations for those models with patterns shifted the most increased by the greatest margin). Those models with the highest correlations before pattern shifting tend to require minimal shifting to obtain the maximum correlation.

Maxima - As illustrated by the maximum ISDME site values in Figure 6.19 and the kriged patterns of Figure 6.16, the predicted and observed maximum values are located within 150 km of the southwestern corner of Pennsylvania. The good agreement is another indication of the better model performances in the spring compared to those of the summer.

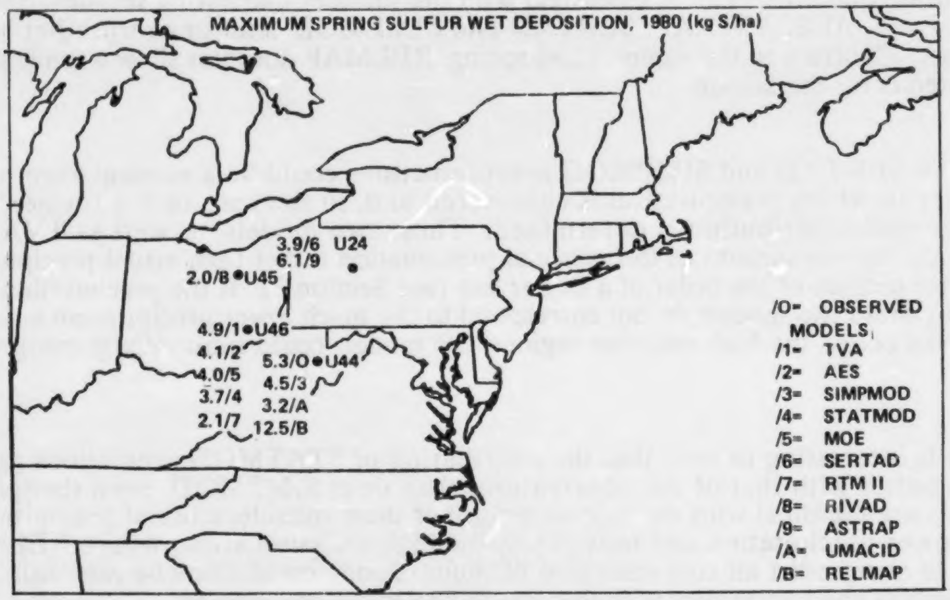


Figure 6.19 **Spring 1980** locations and magnitudes (kg S/ha) of the maximum amount of sulfur wet deposition at the ISDME sites.

Considering the uncertainties of the observations, it cannot be stated that any model fails to accurately predict the location of the maximum value. Therefore, all models pass the criterion for the location of the maximum value.

Both approaches comparing the magnitudes indicate that for eight models the magnitude of the predicted maximum is within the criterion value (30%) of the observed maximum. RELMAP overpredicts by over 50%, while RIVAD and RTM underpredict the magnitude by more than 30%.

Temporal Pattern - The seasonal contribution of only ASTRAP, MOE, RELMAP, RIVAD, RTM and TVA is within 15% of the observed spring contribution of 29%. As is the case for summer, SIMPMOD differs the most from the observed contribution and has a lower seasonal contribution than observed.

6.3.3 Autumn 1980 Results

6.3.3.1 Summary of Distributional Statistics

Figure 6.20 shows the tendency of RIVAD and RTM to underpredict sulfur wet deposition amounts. This is consistent with the summer and spring seasons. On the other hand, AES, ASTRAP, SERTAD and SIMPMOD tend to overpredict in the autumn. Contrary to the summer and spring, RELMAP does not show a tendency to overpredict for this season.

The SERTAD and SIMPMOD overpredictions could be a consequence of the manner in which precipitation is considered and, in the case of the former, how background contribution is determined. These two models, as well as TVA, use statistical representations of the seasonal precipitation rather than actual precipitation rates for periods of the order of a day or less (see Section 2). If the precipitation data used by these two models do not correspond to the much lower precipitation amounts observed across the high emission regions, the models could substantially overpredict there.

It is interesting to note that the distribution of STATMOD predictions agrees much better with that of the observations than does SIMPMOD, even though the models are identical with the sole exception of their consideration of precipitation. The lower precipitation amounts across the high emission areas, where RELMAP tends to overpredict air concentrations of sulfur dioxide could alone be responsible for this model's good agreement with the distribution of the observations.

Figures 6.21 - 6.23 indicate, as those for summer and spring, that SERTAD sulfate air concentrations are much higher than those predicted by other models. RIVAD and

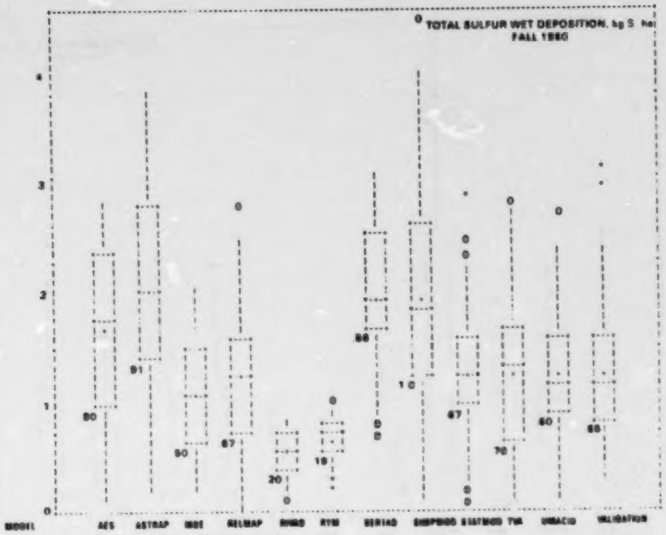


Figure 6.20 **Autumn 1980** box plots of the predictions and observations of sulfur wet deposition (kg S/ha).

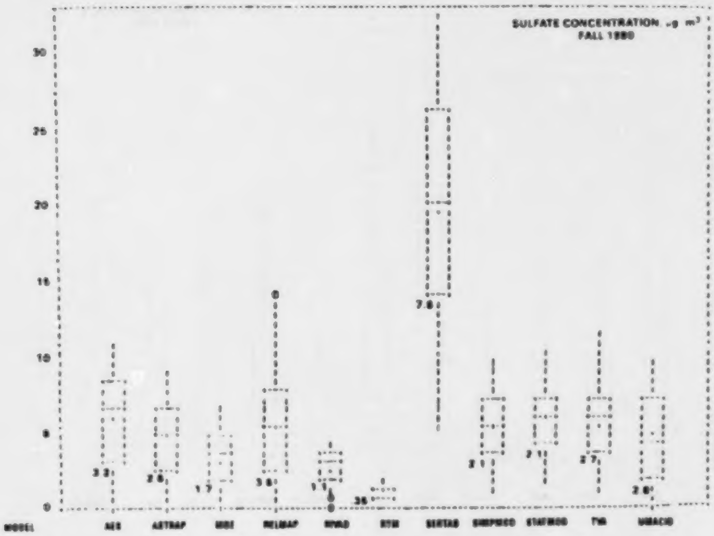


Figure 6.21 **Autumn 1980** box plots of the predictions of sulfate air concentrations (μg/m^3).

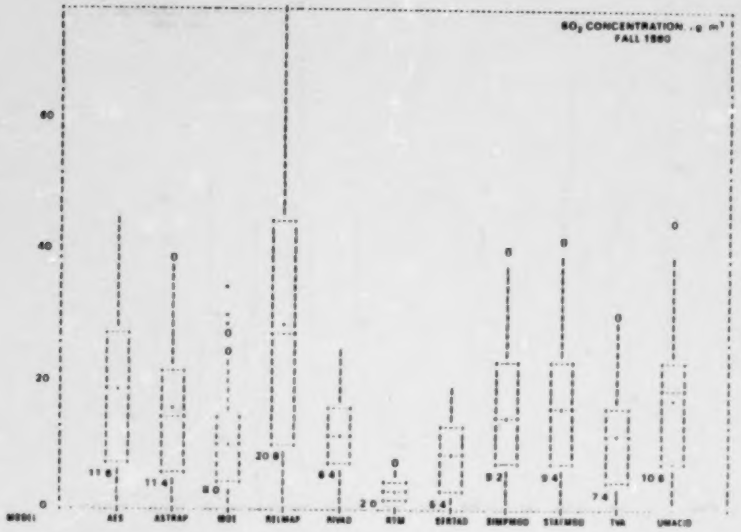


Figure 6.22 **Autumn 1980** box plots of the predictions of sulfur dioxide air concentrations (µg/m³).

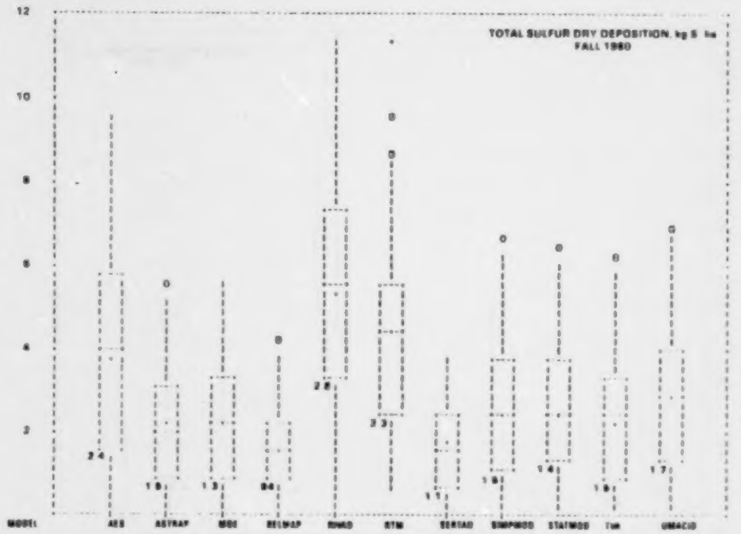


Figure 6.23 **Autumn 1980** box plots of the predictions of sulfur dry deposition (kg S/ha).

RTM sulfur dry deposition amounts are much greater, while their air concentrations of sulfur dioxide and sulfate are much lower than those of other models. This partially explains why these models underpredict sulfur wet deposition amounts. RELMAP air concentrations of sulfur dioxide again are much higher than other predictions.

6.3.3.2 Comparison of Predictions and APN Measurements

In general, with the exception of the statistical models, predictions of sulfate air concentrations are within a factor of two of the measurements at the APN sites (Figure 6.24). Similar to summer and spring, the four statistical models tend to both overpredict and predict the highest sulfate air concentrations. SERTAD, which once again overpredicts by the greatest margin, overpredicts by a factor of ten at the two APN sites closest to the major source regions. Also similar to other seasons, RTM has a tendency to underpredict the sulfate air concentrations by 50%.

With few exceptions, the predictions of sulfur dioxide air concentrations are within a factor of two of the measurements. Despite the doubling of the measured sulfur dioxide air concentrations for autumn (compared to summer and spring), the RTM predictions for autumn are very similar to the RTM predictions for the other two seasons. Consequently, its predictions are low by as much as a factor of four.

As in the case for every season, RELMAP predictions of sulfur dioxide air concentrations at C04 and C05 are two to three times greater than those measured. Also, for all models but SERTAD and MOE, the autumn predictions at C05, the Nova Scotia site, are much higher than the measurements. SERTAD and MOE predictions at this site are very close to the measurements for every season.

6.3.3.3 Spatial Analyses

The observed pattern of sulfur wet deposition for autumn (Figure 6.25) is quite different than those for summer and spring. The pattern for this season shows two distinct and distant areas of maximum values. Both are in Canada, one across southwestern Ontario and the other across the Maritime provinces. In summer and spring, the maximum observed values are typically found across Ohio, West Virginia and western Pennsylvania.

Across the United States the values are relatively low compared to those in Canada. Moreover, the gradients are much flatter than those for summer and spring. Both of these features seem to reflect the shifting of the precipitation pattern northward. Rather than the high precipitation amounts being over the major source region of the United States (as is the case for summer and spring), the higher precipitation amounts in autumn are located across Canada (see Figures B.4 - B.6).

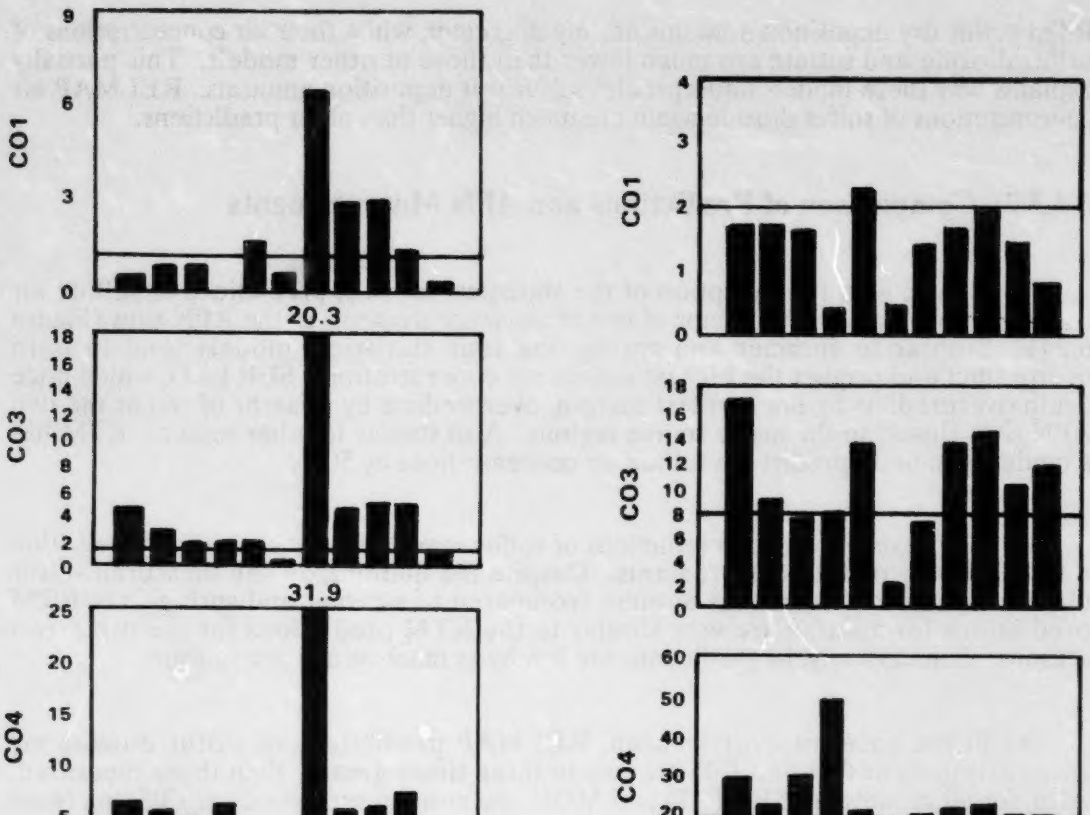


Figure 6.24 Comparison of the **autumn 1980** observations and predictions of the air concentrations (μg/m^3) of sulfur dioxide and sulfate at the APN sites.

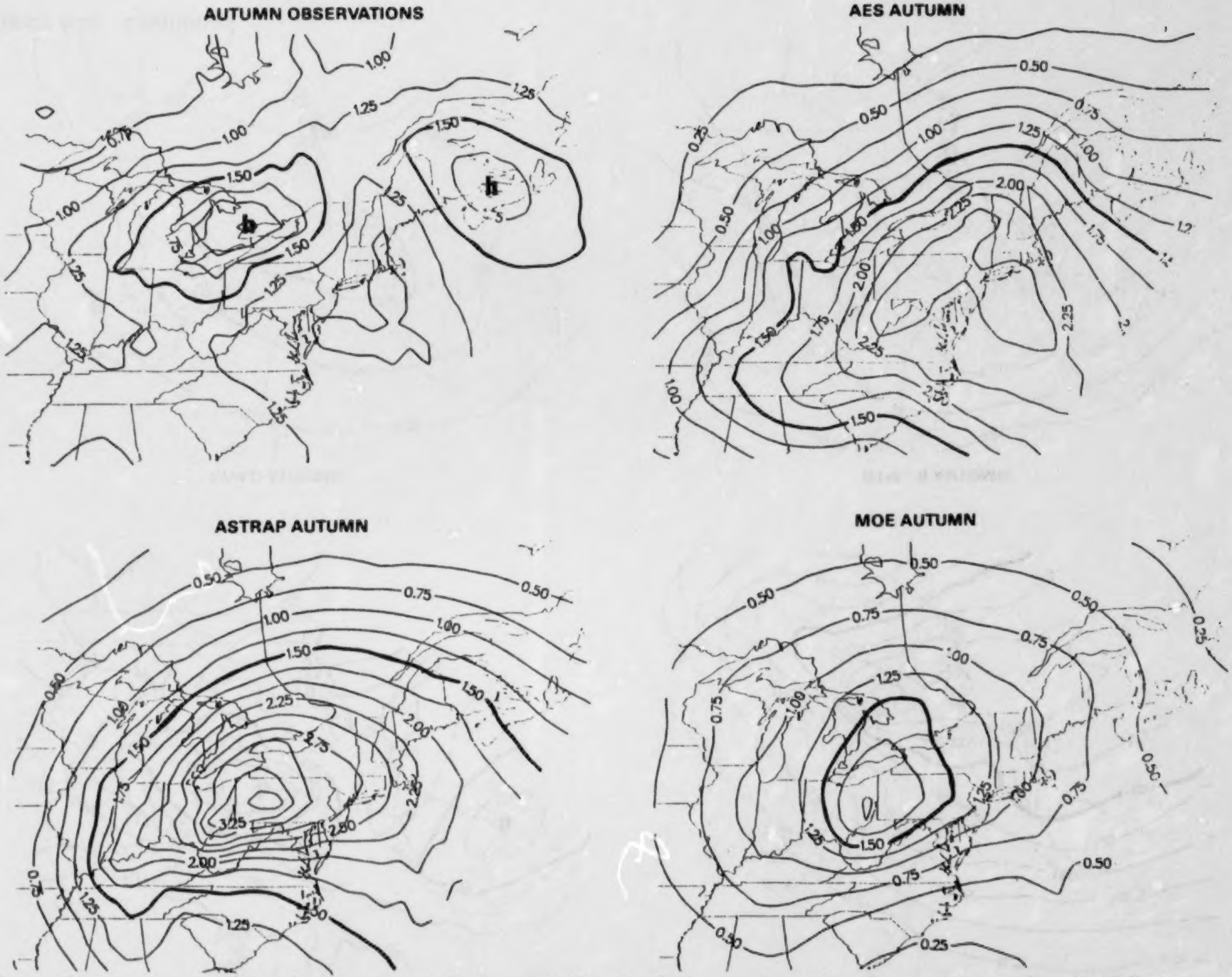
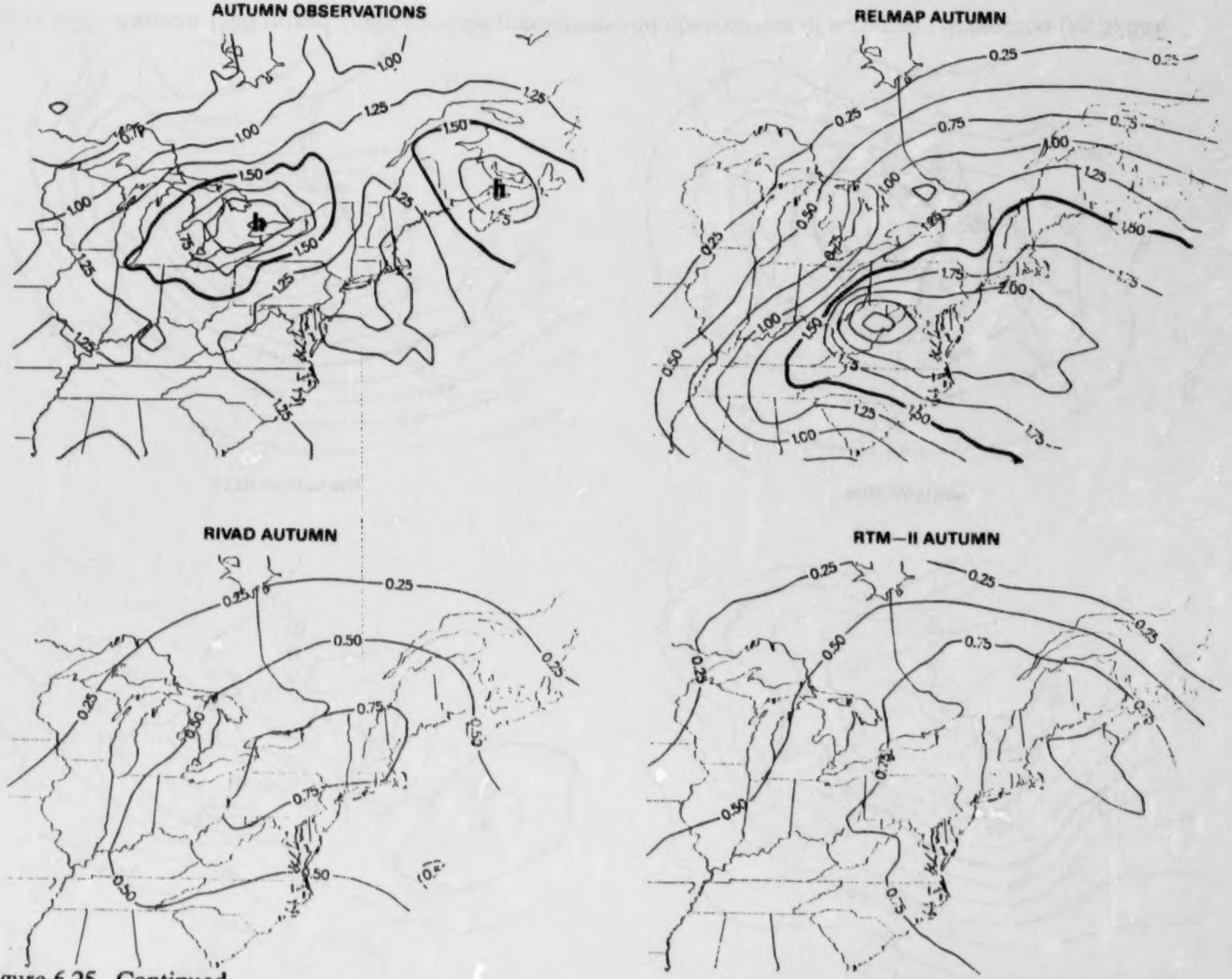


Figure 6.25 **Autumn 1980** kriged analyses of the predictions and observations of sulfur wet deposition (kg S/ha).



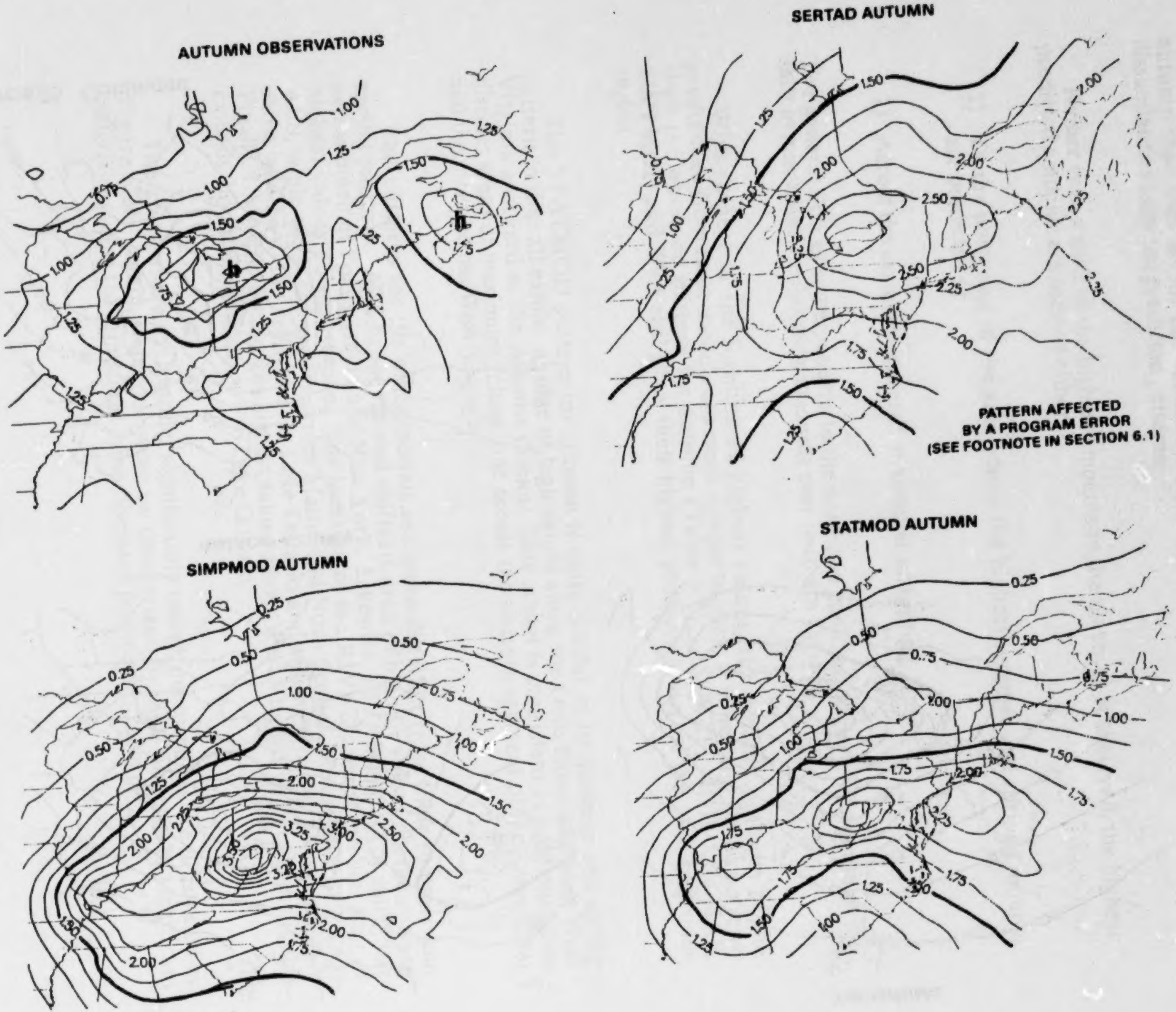
AUTUMN OBSERVATIONS

RELMAP AUTUMN

RIVAD AUTUMN

RTM—II AUTUMN

Figure 6.25 Continued


AUTUMN OBSERVATIONS

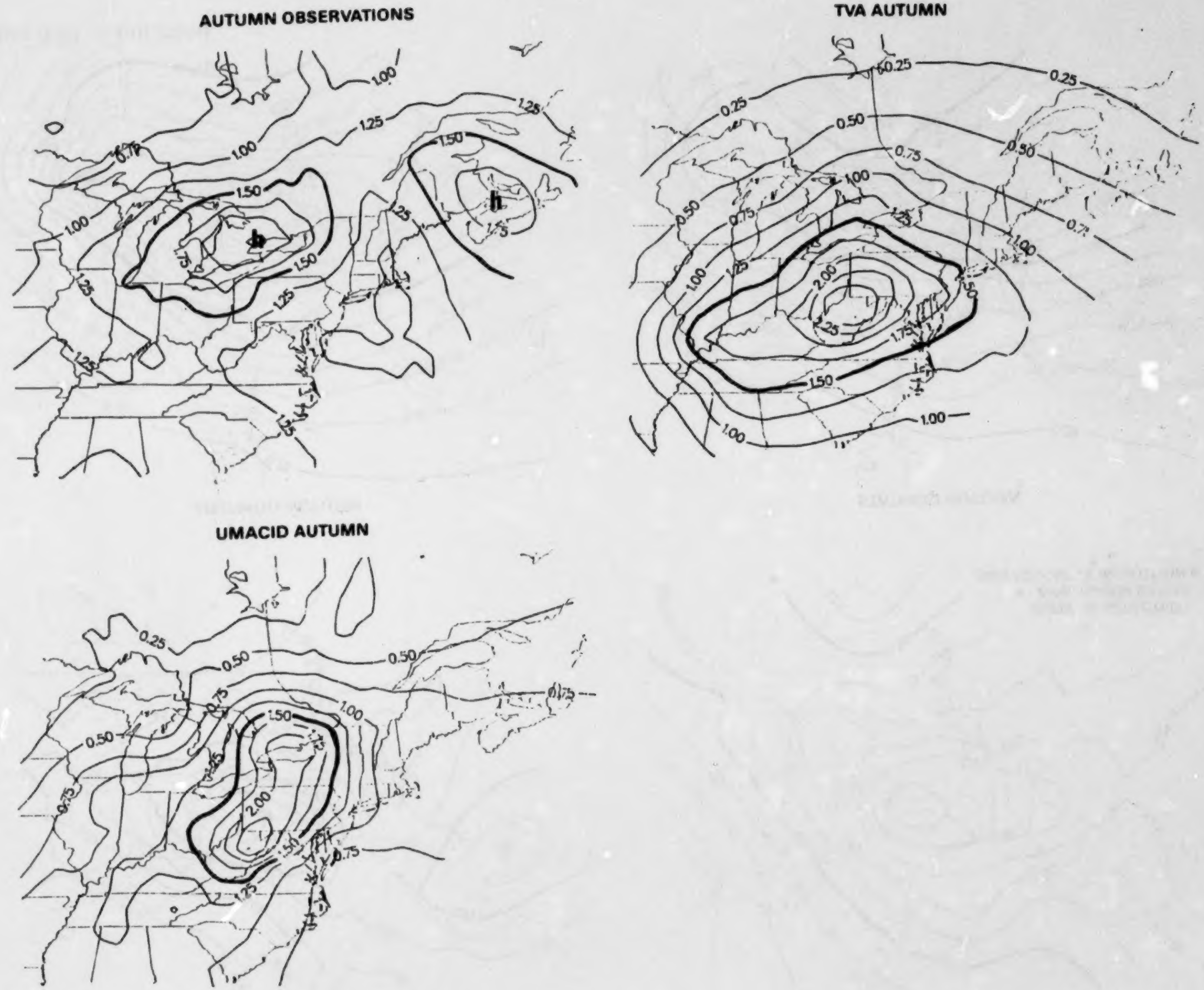
SERTAD AUTUMN

PATTERN AFFECTED BY A PROGRAM ERROR (SEE FOOTNOTE IN SECTION 6.1)

SIMPMOD AUTUMN

STATMOD AUTUMN

Figure 6.25 Continued



AUTUMN OBSERVATIONS

TVA AUTUMN

UMACID AUTUMN

Figure 6.25 Continued

In general, the predicted patterns differ from the observed patterns more for the autumn than they do for the summer or spring. Also, for autumn there are more dissimilarities between predicted patterns.

Rather than a shift of the highest amounts to the north, as is observed, the highest predicted amounts are located either

1) To the northeast of the area where the highest summer and spring amounts are predicted or

2) Across the same general area as summer and spring.

The general shift of some patterns to the northeast could be, in part, a response to the shift in maximum precipitation amounts over southern Canada.

With respect to the location of highest values, the autumn patterns of the predictions from most models are more similar to their summer and spring patterns than the observed pattern for autumn (Table 6.6). ASTRAP, MOE, RELMAP, SIMPMOD and TVA still show their highest predictions across the high emissions region.

The STATMOD pattern for autumn is quite similar to its summer and spring pattern in that all exhibit a ridge of high values along the Ohio River and from West Virginia eastward to the Atlantic Ocean. This model is consistent in predicting two distinct areas of maximum values, one across the western segment of the Ohio River and the other across West Virginia.

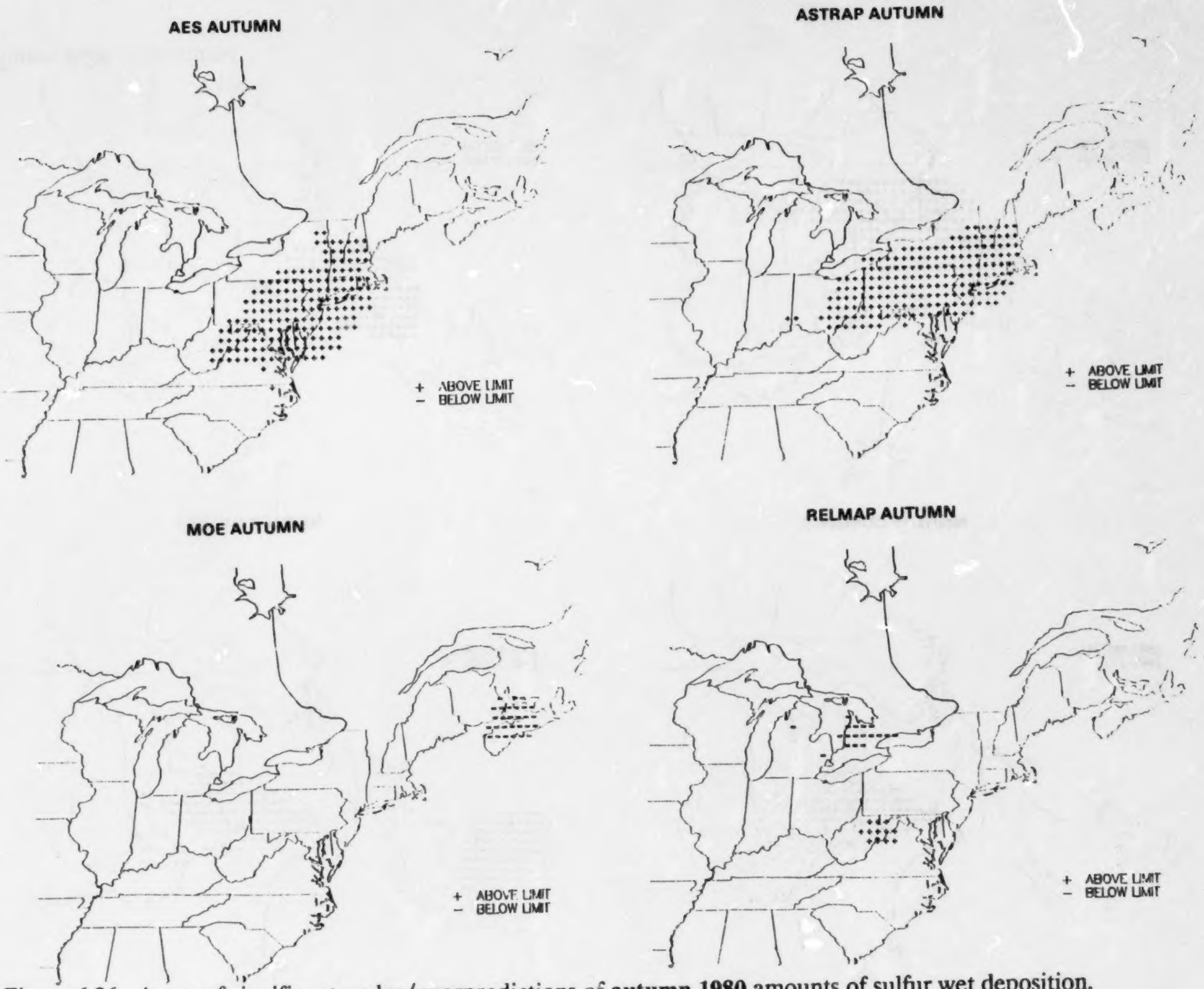
The patterns for the other models are quite different from their summer and spring patterns. AES broadens and shifts its area of highest values northeastward across much of Pennsylvania and New York. Likewise, RIVAD and RTM shift their highest values northeastward. The location of the RIVAD area resembles that of AES, while the RTM area is across Maine and Nova Scotia. Meanwhile, SERTAD shifts its area to the north, across Lake Ontario and eastward to the Atlantic Ocean. Finally, UMACID predicts its highest values across West Virginia, as before, but it also extends this area northeastward to Lake Ontario.

The areas where each model significantly under/overpredicts are illustrated by Figure 6.26. Again, the models can be categorized by their patterns of significant differences using the same three types discussed previously (Table 6.7).

TABLE 6.6. COMPARISON OF THE GEOGRAPHICAL REGIONS WHERE EACH MODEL PREDICTS THE HIGHEST AMOUNTS OF SULFUR WET DEPOSITION FOR SUMMER/SPRING VERSUS AUTUMN 1980.

MODEL	SUMMER/SPRING	AUTUMN
OBSERVED	Ohio, Pennsylvania, West Virginia	SW Ontario, Maritime Provinces
AES	Ohio, Pennsylvania, West Virginia	West Virginia to N New York
ASTRAP	W Pennsylvania	no difference
MOE	Ohio, W Pennsylvania, Lake Erie	no difference
RELMAP	West Virginia	no difference
RIVAD	Ohio, Pennsylvania, West Virginia	Pennsylvania to New Hampshire
RTM	Ohio, Pennsylvania, West Virginia	Maine, Nova Scotia
SERTAD*	Ohio, Pennsylvania, Lake Erie	Lake Ontario
SIMPMOD	Pennsylvania, West Virginia, W Maryland	no difference
STATMOD	Ohio, West Virginia	West Virginia to the Atlantic Ocean
TVA	Ohio, Pennsylvania, West Virginia	no difference
UMACID	Ohio, Pennsylvania, West Virginia	West Virginia to Lake Ontario

* See footnote in Section 6.1.



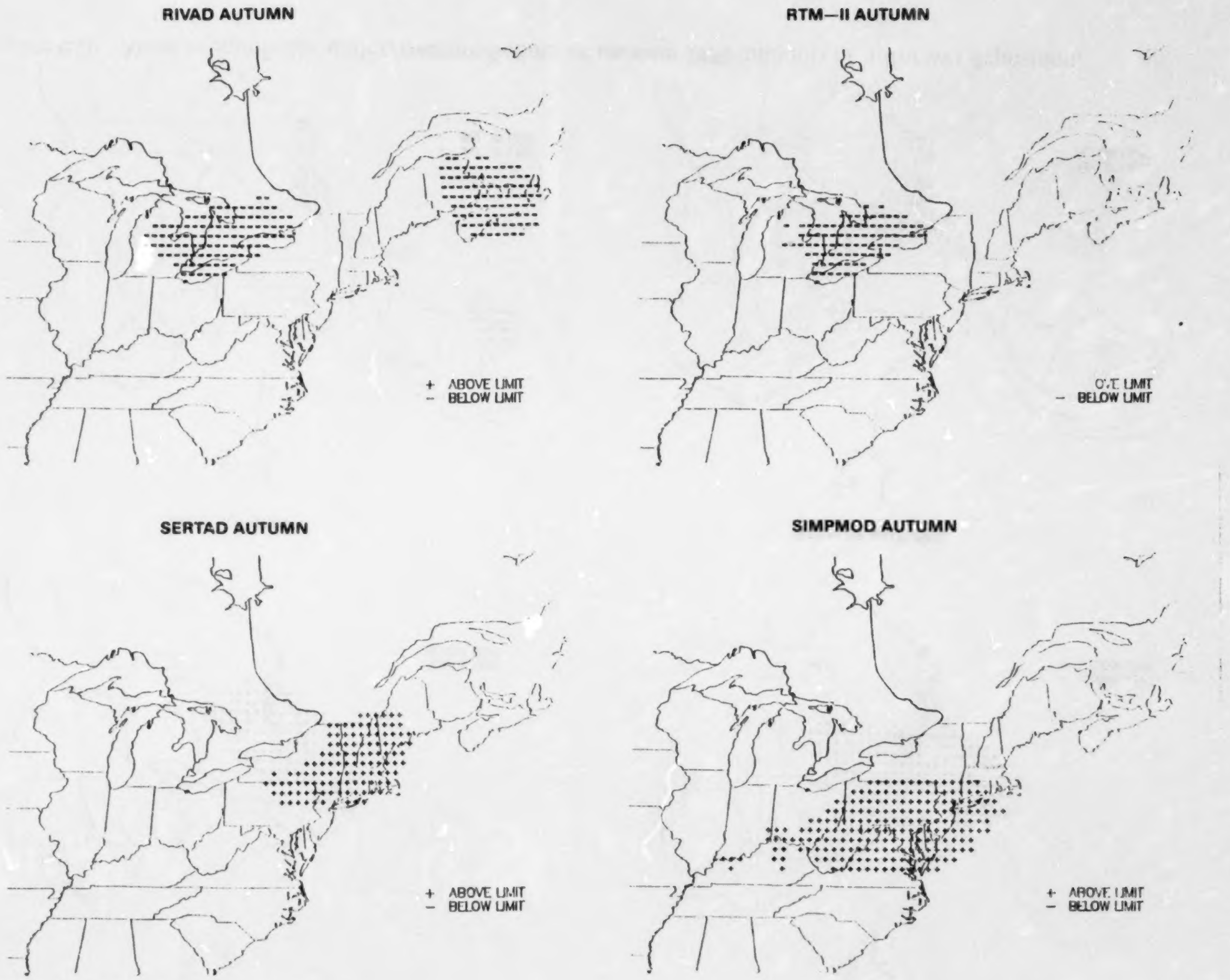
AES AUTUMN

ASTRAP AUTUMN

MOE AUTUMN

RELMAP AUTUMN

Figure 6.26 Areas of significant under/overpredictions of **autumn 1980** amounts of sulfur wet deposition.



RIVAD AUTUMN

RTM—II AUTUMN

SERTAD AUTUMN

SIMPMOD AUTUMN

Figure 6.26 Continued

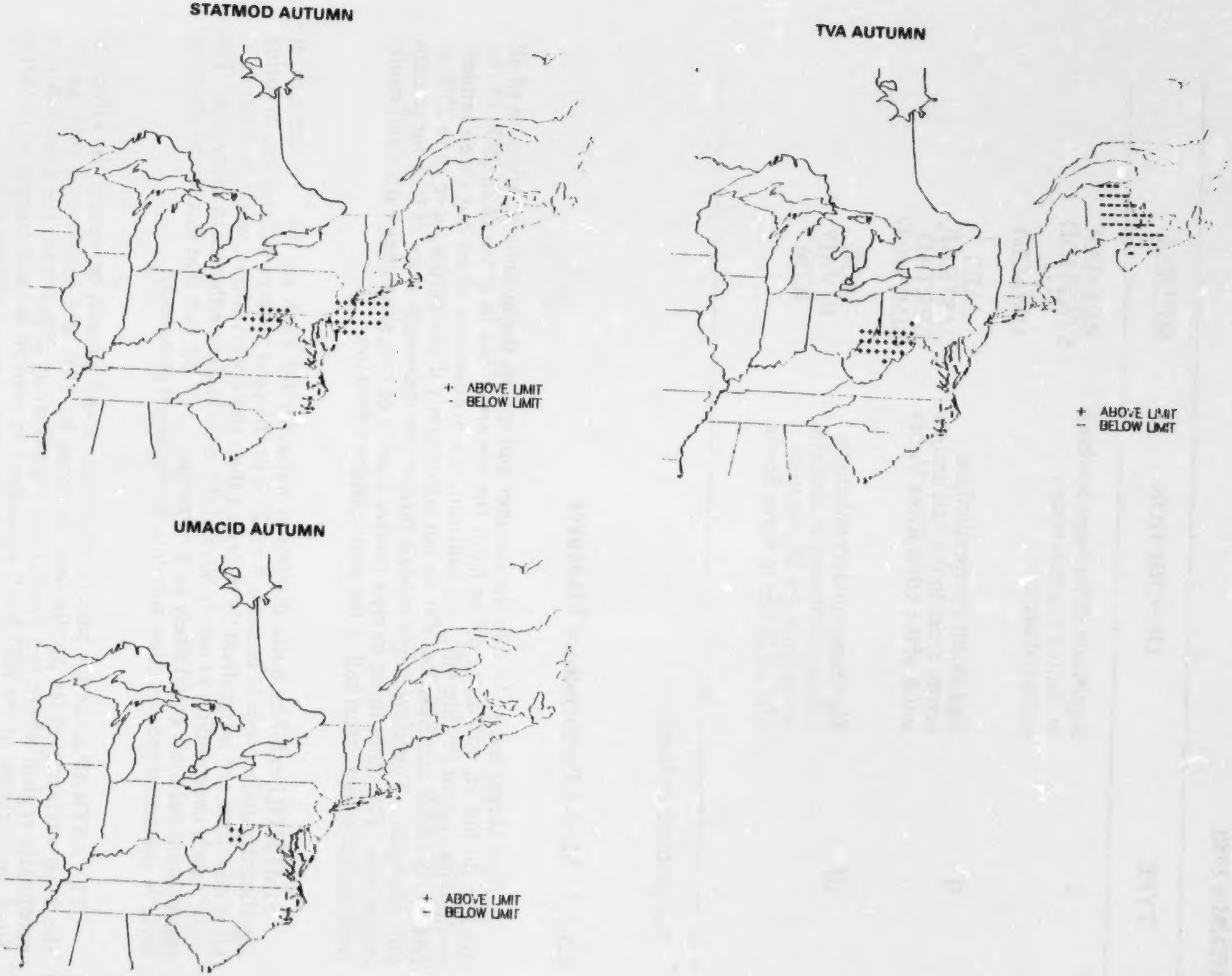


Figure 6.26 Continued

TABLE 6.7 TYPES OF PATTERNS OF SIGNIFICANT DIFFERENCES FOR AUTUMN 1980.

TYPE	DESCRIPTION	MODELS
I	Significant differences nowhere or limited to small widely scattered areas	RELMAP STATMOD TVA UMACID
II	Significant overpredictions across areas to the east and/or south of the Ohio River Valley	AES ASTRAP* SERTAD* SIMPMOD
III	Significant underpredictions across continuous or sporatic areas from the Midwestern United States to Nova Scotia	MOE RIVAD RTM

* See footnote in Section 6.1.

6.3.3.4 Model Performance Measures

Pattern Magnitude - Unlike the summer and spring, the autumn predictions of all models do not significantly differ from the observations in more than 20% of the composite region (Table 6.8). In addition, the predictions of all but three models (AES, ASTRAP and SIMPMOD) do not significantly differ across more than 25% of any subregion. Therefore, eight models receive the maximum score of four for pattern magnitude. The other three models receive a score of two, since they each significantly overpredict in more than half of the southeastern subregion.

SIMPMOD exhibits quite dissimilar behavior for each season. In summer, it significantly underpredicts in more than 29% of the cells in three subregions, in spring all predictions are insignificantly different than the observations, while in autumn, it significantly overpredicts in nearly 80% of the cells in the southeastern subregion. The inconsistent behavior could likely be a consequence of the fact that this model does not directly use actual precipitation amounts, as explained previously.

The ASTRAP score for pattern magnitude could actually be zero if the effects of the high variance of its predictions increase by 25% the number of cells in the composite region where the predictions significantly differ from the observations. Since the variance of the RELMAP predictions in autumn is small compared to that of the observations, its score would not likely change as a result of these effects.

TABLE 6.8 SUMMARY OF MODEL PERFORMANCE MEASURES FOR AUTUMN 1980.

	AES	AST	MOE	REL	RIV	RTM	SER	SIM	STA	TVA	UMA	OBS
I. Spatial Pattern Replication												
A. Magnitude												
1. Area of Significant Differences (%)*												
a. Northeast	1.5+	1.5+	9.8-	0.0	21.4-	0.0	10.5+	0.0	0.0	13.2-	0.0	
b. North Central	9.5+	14.9+	0.0	4.1-	23.8-	18.8-	14.0+	3.2+	0.2-	0.0	0.0	
c. Southwest	0.0	11.0+	0.0	0.0	0.0	0.0	0.0	18.4+	0.0	2.6+	0.0	
d. Southeast	63.8+	53.9+	0.0	7.9+	0.0	0.0	13.8+	78.9+	14.5+	10.5+	2.6+	
e. All 4 regions	13.1+	16.3+	2.4-	2.8	14.9-	7.6-	10.2+	16.2+	2.1+	5.2	0.4+	
2. Median (kg S/ha)	1.92	2.24	1.15	1.36	0.59	0.76	2.04	2.03	1.33	1.40	1.17	1.28
(p-value)	0.006	0.000	0.254	0.649	0.000	0.000	0.000	0.000	0.649	0.649	0.362	
3. Dispersion (kg S/ha)	0.46	0.62	0.22	0.38	0.03	0.03	0.28	0.86	0.37	0.40	0.36	0.41
(p-value)	0.353	0.097	0.029	0.412	0.000	0.000	0.127	0.011	0.408	0.477	0.349	
4. Bias-High (kg S/ha)	-0.15	-1.02	0.13	0.38	1.04	1.00	-0.54	-0.76	0.09	-0.06	-0.01	
(p-value)	0.24	0.00	0.25	0.07	0.00	0.00	0.00	0.02	0.37	0.42	0.49	
5. Bias-Low (kg S/ha)	-0.46	-0.56	0.24	-0.10	0.59	0.42	-0.76	-0.50	0.03	0.09	0.21	
(p-value)	0.01	0.00	0.03	0.26	0.00	0.00	0.00	0.02	0.42	0.29	0.10	
6. Mean Squared Error(kg S/ha)^2	1.03	1.39	0.51	0.76	0.98	0.80	0.81	1.76	0.79	0.82	0.46	
B. Position												
1. Centroids (km)	139	65	71	182	71	97	65	126	147	99	38	
2. Major Axis (degrees)**	+7.3	+16.6	+18.2	+3.3	+12.2	+8.6	+13.3	+20.7	+23.2	+21.0	+8.7	29.0
3. Shift (km)	660	440	330	592	347	934	246	454	592	440	453	
C. Shape												
1. Correlation-No Shift	-0.39	0.37	0.59	-0.58	0.17	-0.19	0.54	-0.43	-0.50	-0.26	0.30	
2. Correlation-With Shift	0.55	0.81	0.86	0.61	0.78	0.20	0.71	0.83	0.74	0.89	0.88	
II. Maxima												
A. Location-Kriged (km)	338	183	164	338	315	135	0	338	285	338	338	
B. Magnitude-Kriged (%)**	0	0	0	0	-52	-42	0	0	0	0	0	
Magnitude-Site (%)**	0	+1	-24	0	-68	-61	0	+18	0	0	0	
III. Temporal Pattern Replication												
A. Seasonal Contribution (%)	19.4	19.1	17.5	11.7	13.4	15.5	22.2	22.3	22.2	17.3	20.7	16.0
(Percent Difference)**	+21	+19	+9	-27	-16	-3	+39	+39	+39	+8	+29	

Underscored measures are primary performance measures.

* "+" and "-" denote the prevalence of overpredictions and underpredictions, respectively.

** Values listed for each model represent the value of the prediction minus the value of the observation. Percent differences are normalized by observed values.

Based on the descriptive statistics alone, which do not consider the uncertainties of the observations, five of the eight models receiving a score of four (MOE, RELMAP, STATMOD, TVA and UMACID) could be identified as better performers. However, RIVAD, RTM and SERTAD could receive a lower score if the descriptive statistics were used alone. Those models receiving a score of two show the poorest results based on the descriptive statistics.

As mentioned previously, a large percentage of the SERTAD predictions of sulfur wet deposition are attributed to a background contribution. As an example, about 50% of the magnitude of the SERTAD predictions across Ohio are attributed to its background contribution. This is based on a typical autumn precipitation amount for Ohio of 120 mm (Figure B.6).

Pattern Position - Although the patterns of predictions visually appear to differ from the pattern of observations for autumn (Figure 6.25), most models receive the maximum score for pattern magnitude. This can be explained by the small range in the observed values compared to the uncertainty of the observations. Even though the positioning of the patterns is obviously different, much of the time, the predictions are within the uncertainty envelope. The measures for pattern position, however, quantify the pattern differences, underscoring the necessity for a balanced and comprehensive set of model evaluation measures.

Only five models (ASTRAP, MOE, RIVAD, SERTAD and UMACID) have centroids within the criterion value of 75 km from the observed centroid (Figure 6.27). The centroids for the other models are as much as 180 km southeast of the observed centroid.

The difference between the orientation of the major axes of the predictions and observations does not exceed 10 degrees, the criterion value, for only four models (AES, RELMAP, RTM and UMACID). The orientation differs the most (in excess of 20 degrees) for three of the four statistical models (SIMPMOD, STATMOD and TVA). (As discussed previously, the attributes of the pattern for the other statistical model, SERTAD, can be dominated by the precipitation pattern). The orientation of the major axes for each model is rotated clockwise from the observed major axis. This is consistent with the summer, but not the spring, results.

The differences in pattern positioning is apparent from the distances the predicted patterns are shifted to maximize the correlation. In contrast to summer and spring, the shifts in the autumn patterns exceed 400 km for all but three models, while for the other two seasons, the shift distances exceed 250 km for only two or three models.

Pattern Shape - In contrast to summer and spring, and in concuurence with the previously discussed values of the measures for autumn, the correlations for autumn, ranging from -0.6 to 0.6 (Table 6.8), reflect the relatively poor agreement between the spatial patterns of the predictions and observations. For six models (AES, RELMAP, RTM, SIMPMOD, STATMOD and TVA) the correlations are negative, indicating that these models tend to predict their highest values where the lowest values are

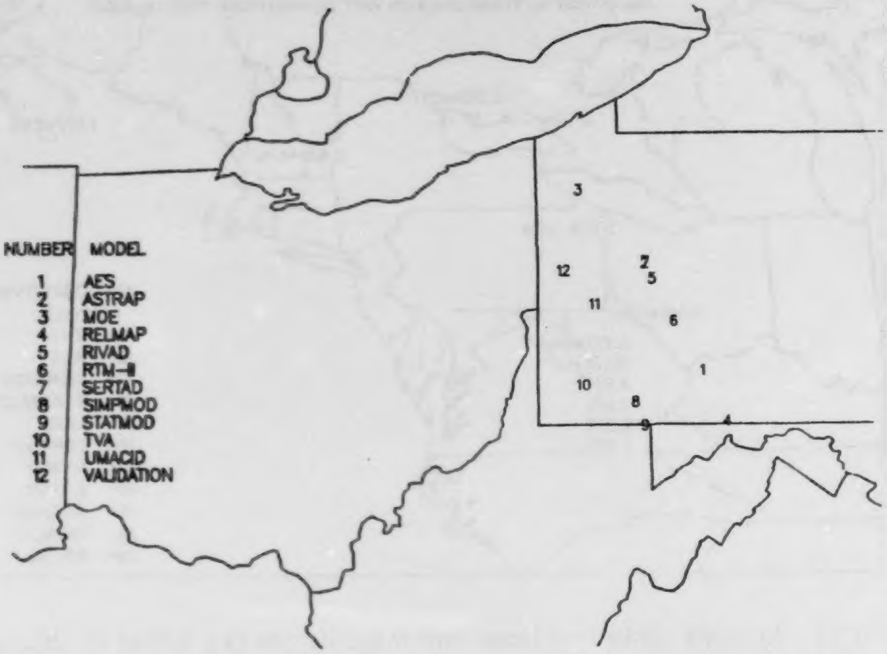


Figure 6.27 **Autumn 1980** locations of the centroids.

observed and vice versa.

Consistent with the summer and spring, the MOE and SERTAD correlations for autumn, although they are less than 0.6, are the highest (in the positive sense) than those of the other models. However, the correlations for UMACID, which has correlations greater than 0.7 for summer and spring, has an autumn correlation of only 0.3.

Pattern shifting dramatically increases the correlations for all cases but one (that for RTM remains less than 0.3). This indicates that the RTM pattern shape is so dissimilar to that of the observations that pattern shifting fails to significantly improve the correlation. With the exception of AES, RELMAP and RTM, at least half the variance is explained by the models after the patterns are shifted.

Maxima - As seen from Figures 6.25 and 6.28, the locations of the maximum values are widely scattered, more so than for any other season. Only two models (RTM and SERTAD) predict their maximum values in Canada. RIVAD locates its maximum prediction in New Hampshire where the observed pattern shows a trough of low values. Most other models locate their maximum predictions in or near West Virginia.

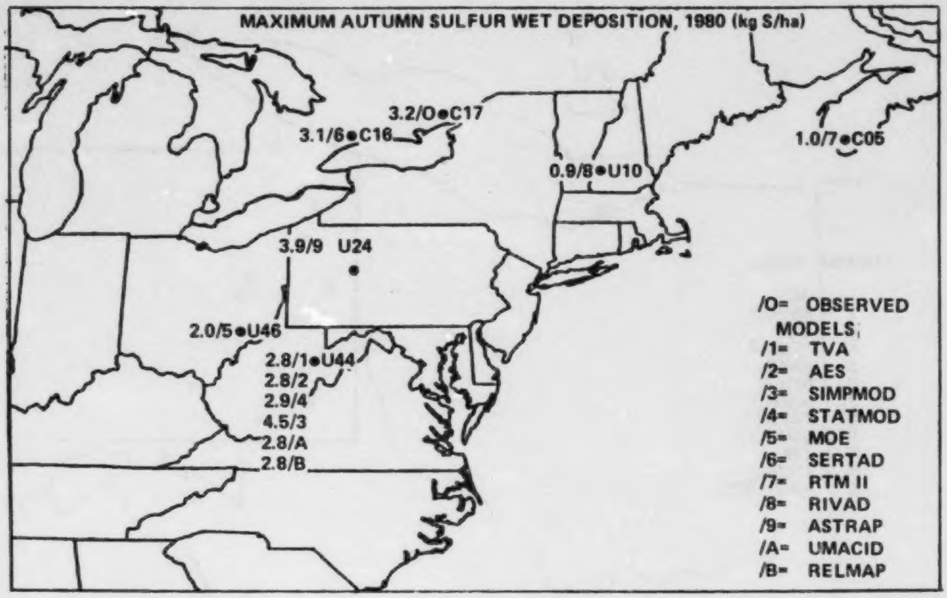


Figure 6.28 **Autumn 1980** locations and magnitudes (kg S/ha) of the maximum amount of sulfur wet deposition at the ISDME sites.

The values of the measure for the location of the maxima (Table 6.8) indicate that only SERTAD passes the criterion that the predicted maximum be within 100 km of the observed maximum. The location of the maximum predictions of the other models is over 280 km from the nearest possible location of the observed maximum.

It is interesting to point out a feature of the kriging approach to compute the distance between the predicted and observed maximum. The observed pattern for autumn (Figure 6.25) shows two distinct and distant (i.e., greater than 1000 km) peaks, the largest one near Toronto and the other one (about 30% lower in magnitude) over Nova Scotia and New Brunswick. Even though the locations of the RTM and SERTAD maxima are over 1000 km apart (the former being in Nova Scotia, while the latter is near Toronto), the kriging approach does not ignore the fact that each location is very near a possible location of the observed maximum (Table 6.8).

With the exception of RIVAD and RTM, the magnitudes of the predicted maxima are within 25% of that observed. Consequently, all but these two models, which underpredict the magnitude by about 50%, pass the criterion for the magnitude of the maximum.

Temporal Pattern - The autumn contribution of the predictions are within 10% of the observed autumn contribution (16% of the annual total) for only three models

(MOE, RTM and TVA). Therefore, only these models pass the criterion. SERTAD, SIMPMOD and STATMOD each overpredict the autumn contribution by nearly 40%.

6.3.4 Winter 1980 Results

6.3.4.1 Summary of Distributional Statistics

As seen from Figure 6.29 AES, ASTRAP and SIMPMOD tend to overpredict, while RIVAD and RTM tend to underpredict both the magnitudes and the standard deviation of the sulfur wet deposition amounts for winter. The second quartile predictions of ASTRAP and SIMPMOD are greater than the third quartile observations.

The winter precipitation pattern (see Figure B.3) is similar to that of autumn in that precipitation amounts across the high emission regions are relatively low. As in the case of autumn, this could partially explain the higher SIMPMOD predictions. SERTAD predictions, which for the autumn were nearly as high as SIMPMOD's, are much lower than SIMPMOD winter predictions.

Figures 6.30 - 6.32 indicate that for the winter, as for the other seasons, SERTAD sulfate air concentrations, RELMAP sulfur dioxide air concentrations and sulfur dry deposition predictions of AES, RIVAD and RTM are substantially higher than those predicted by the other models. Consistent with other seasons, RIVAD and RTM sulfate air concentrations are the lowest. In addition, MOE sulfate air concentrations are relatively low.

6.3.4.2 Comparison of Predictions and APN Measurements

Also like the other seasons, the statistical models tend to both overpredict and predict the highest sulfate air concentrations at the APN sites (Figure 6.33). Again, SERTAD predictions are as much as six times higher than the measurements. With the exception of the predictions of the statistical models and RELMAP predictions at C04, the predictions are within a factor of two of the measurements.

It is interesting to note the small seasonal variations in sulfate air concentrations at these sites. In contrast, the winter sulfur dioxide air concentrations increase by factors ranging from 2.5 at C04 to nearly 6 at both C01 and C03.

With the exception of the RELMAP predictions, there are only six out of a possible forty cases when the predictions of sulfur dioxide air concentrations differ from the observations by more than a factor of two. The tendency for RIVAD to

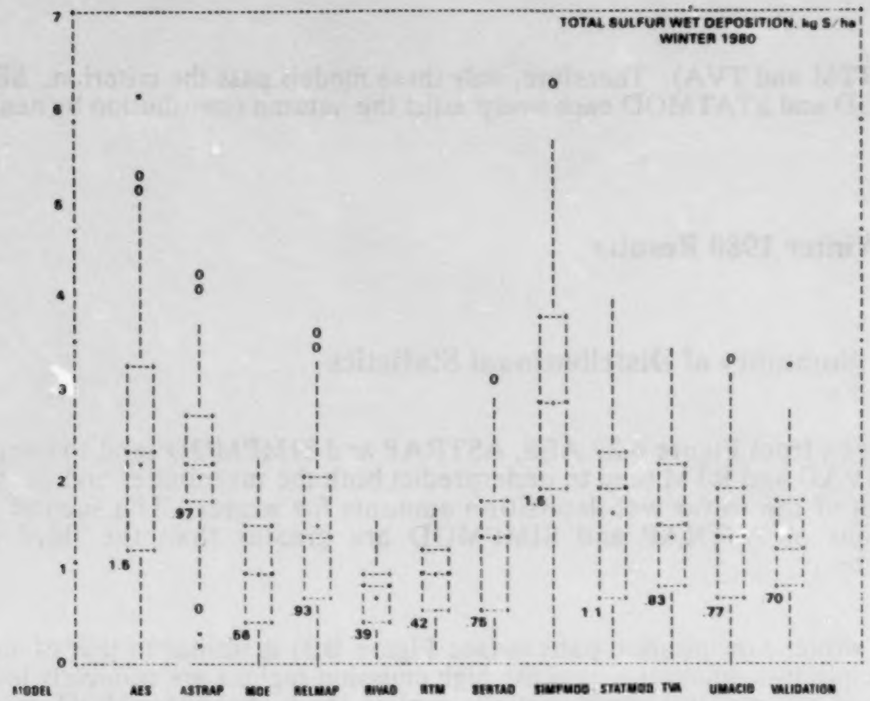


Figure 6.29 Winter 1980 box plots of the predictions and observations of sulfur wet deposition (kg S/ha).

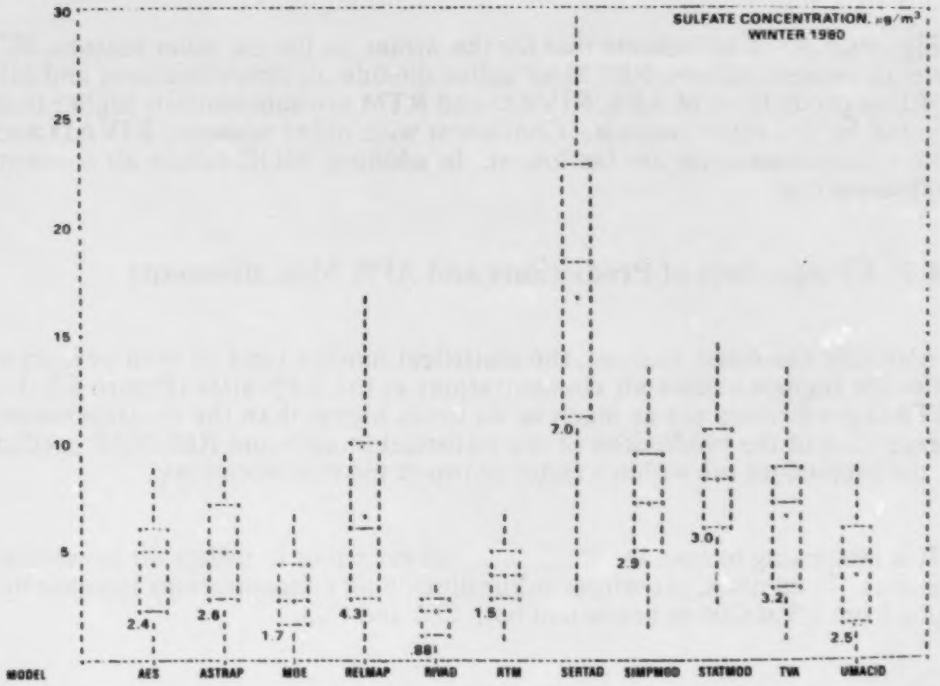


Figure 6.30 Winter 1980 box plots of the predictions of sulfate air concentrations (µg/m³).

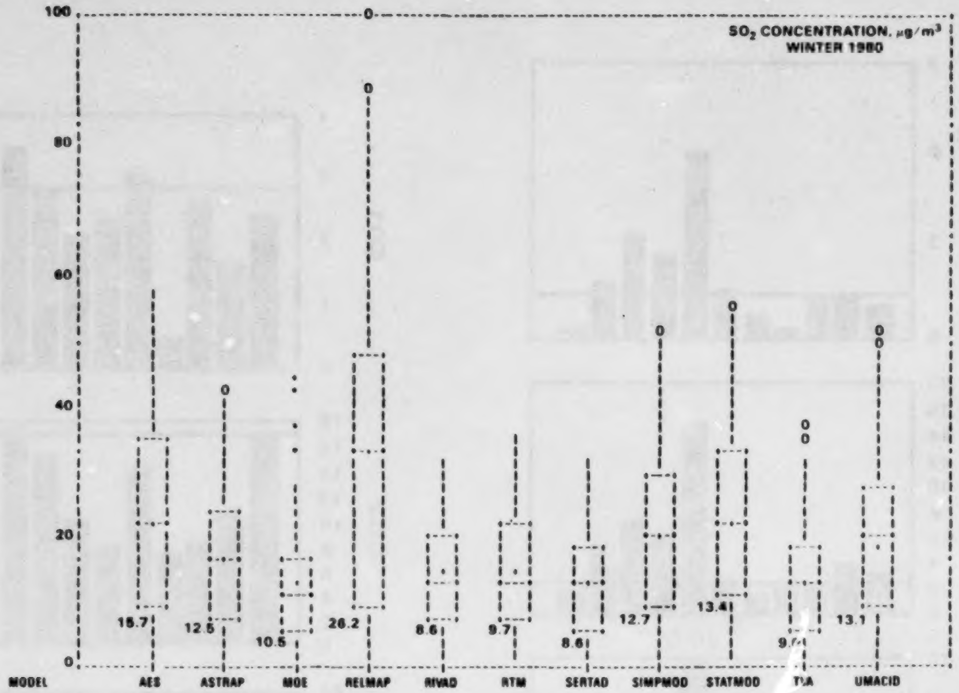


Figure 6.31 **Winter 1980** box plots of the predictions of sulfur dioxide air concentrations (µg/m^3).

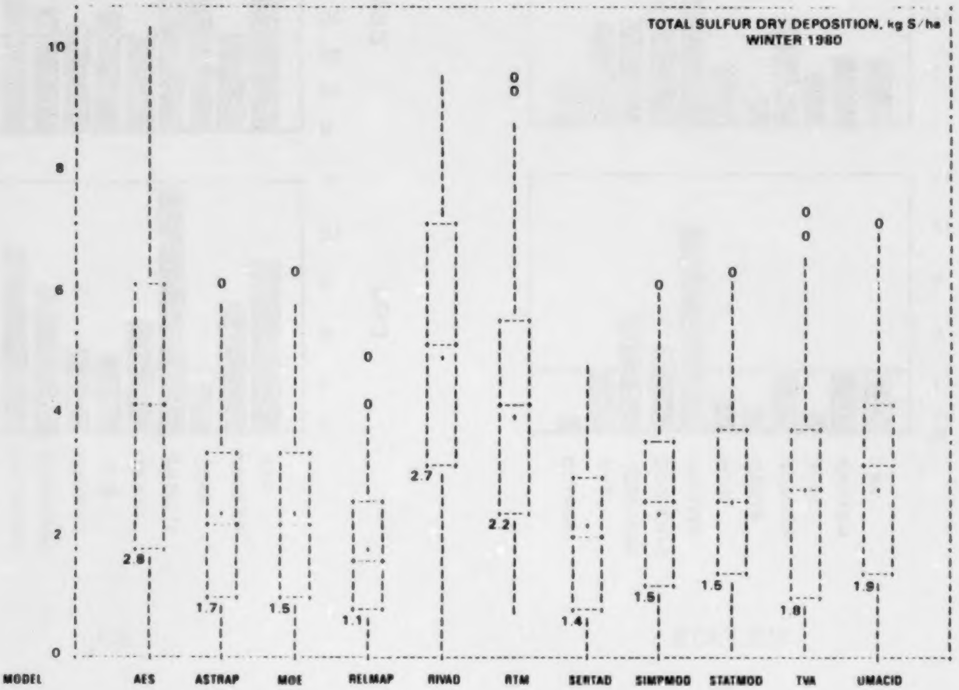


Figure 6.32 **Winter 1980** box plots of the predictions of sulfur dry deposition (kg S/ha).

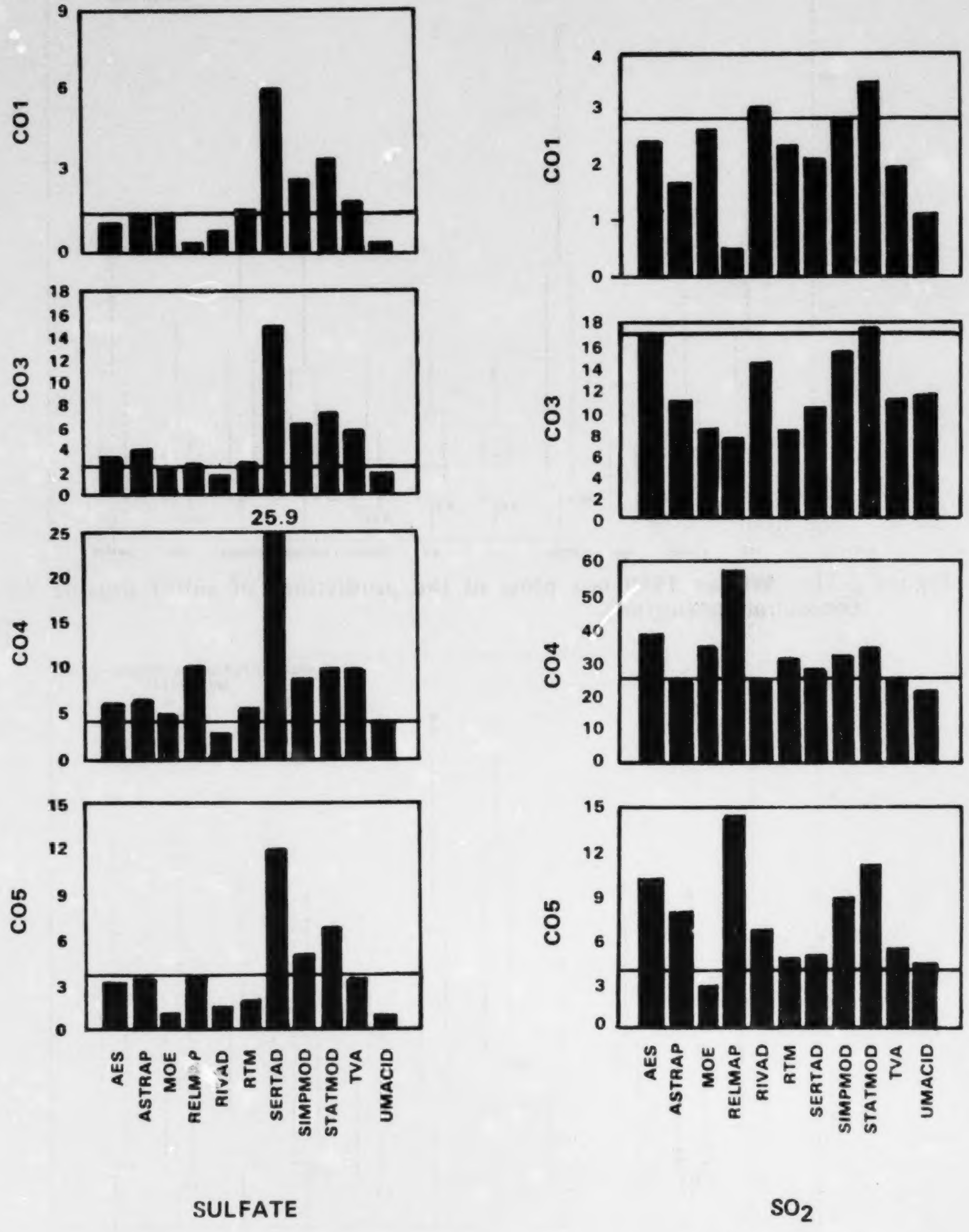


Figure 6.33 Comparison of the **winter 1980** observations and predictions of the air concentrations ($\mu g/m^3$) of sulfur dioxide and sulfate at theAPN sites.

overpredict the sulfur dioxide air concentrations is not apparent at the APN sites in winter, when its predictions are comparable to the measurements. On the other hand, RELMAP predictions at C04 and C05 are still three to four times higher than the measurements.

6.3.4.3 Spatial Analyses

Like the autumn pattern of the observations (Figure 6.25), the pattern of the winter observations (Figure 6.34) shows two distinct and distant areas of maximum values, one of which is located across southern Ontario. However, for winter, the second area is located across the southeastern United States. The southern area of maximum values coincides with the peak amounts of precipitation (40 cm). The winter precipitation amounts to the north of the Ohio River and Maryland are half as great. Minimum values are found across eastern Canada, the northeastern United States and Indiana.

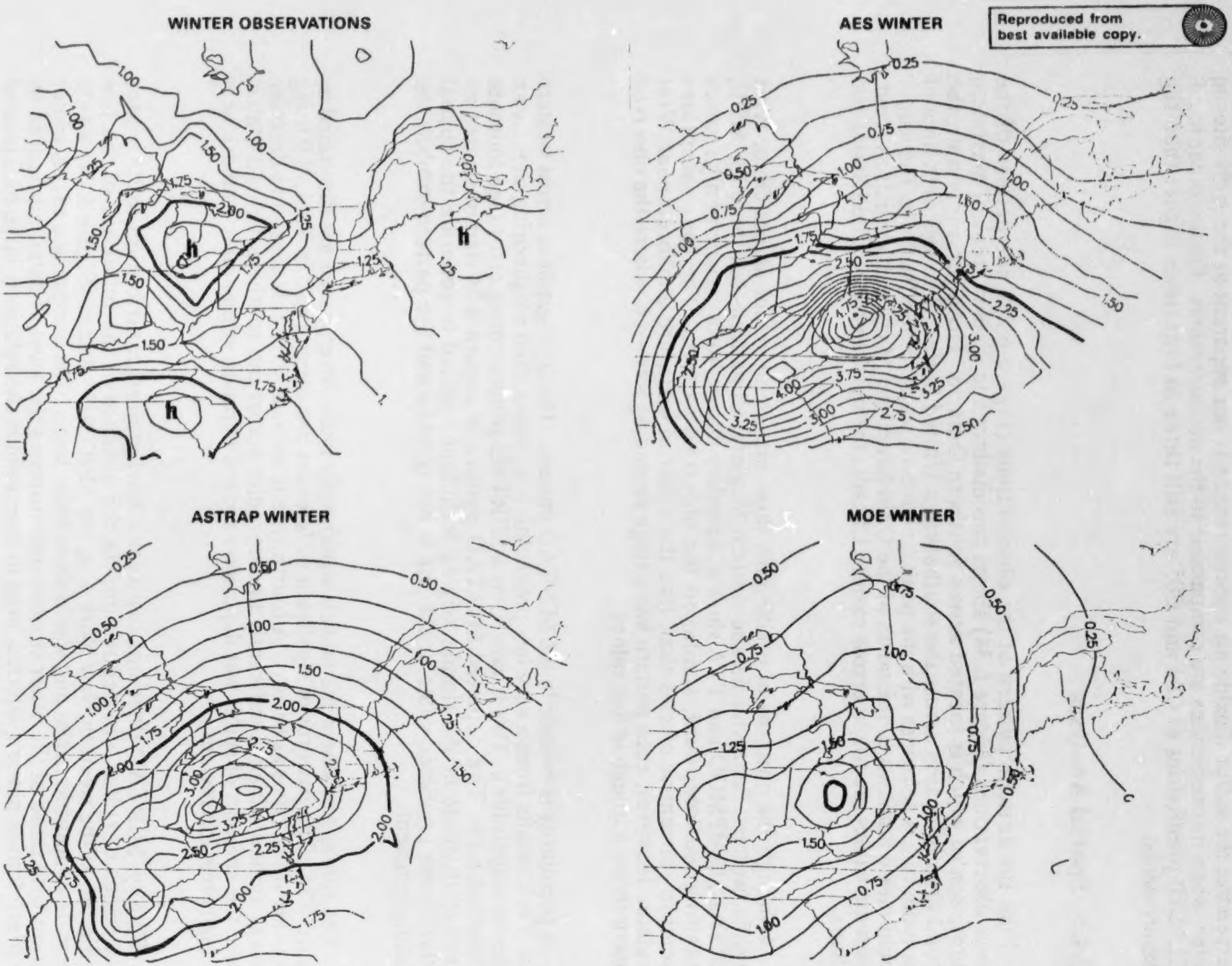
Although the patterns of predictions show maximum values across Ohio, West Virginia, western Pennsylvania and western Virginia, all patterns but those for MOE, SERTAD, SIMPMOD and TVA show a secondary maximum or a ridge of high values extending *southwestward*, similar to the observed pattern. The four exceptions resemble concentric circles that, like the other models, are anchored near West Virginia. However, each pattern has a ridge extending *westward*, where the observed pattern shows a trough of low values.

A peculiarity is evident in the SERTAD pattern. The tight gradient across western New York results from a very low prediction (i.e., lower than neighboring sites by an order of magnitude). This is caused by a SERTAD programming error (see footnote in Section 6.1). The depicted SERTAD pattern in Figure 6.34 demonstrates an intrinsic attribute of the simple kriging technique applied to generate the spatial patterns: the predicted value at the site is not ignored and the pattern exhibits the resulting gradient.

The subregions where the models significantly under/overpredict are illustrated by Figure 6.35. The patterns of signifcant differences are categorized in Table 6.9 using the same three types as before. A fourth type is necessary for this season to describe several models which significantly underpredict across the southern United States, where the amounts of sulfur wet deposition are not normally comparable to those near the high emission regions.

Finally, SERTAD does not consider a background contribution to its winter predictions of sulfur wet deposition. In the absence of the influence of the low prediction at the western New York site, the SERTAD pattern (Figure 6.34) would resemble concentric circles and would show little detail. The comparison of this winter pattern with those of the other seasons demonstrates the level of detail introduced by the precipitation pattern, which is used in determining the relatively large background contributions.

Reproduced from
best available copy.

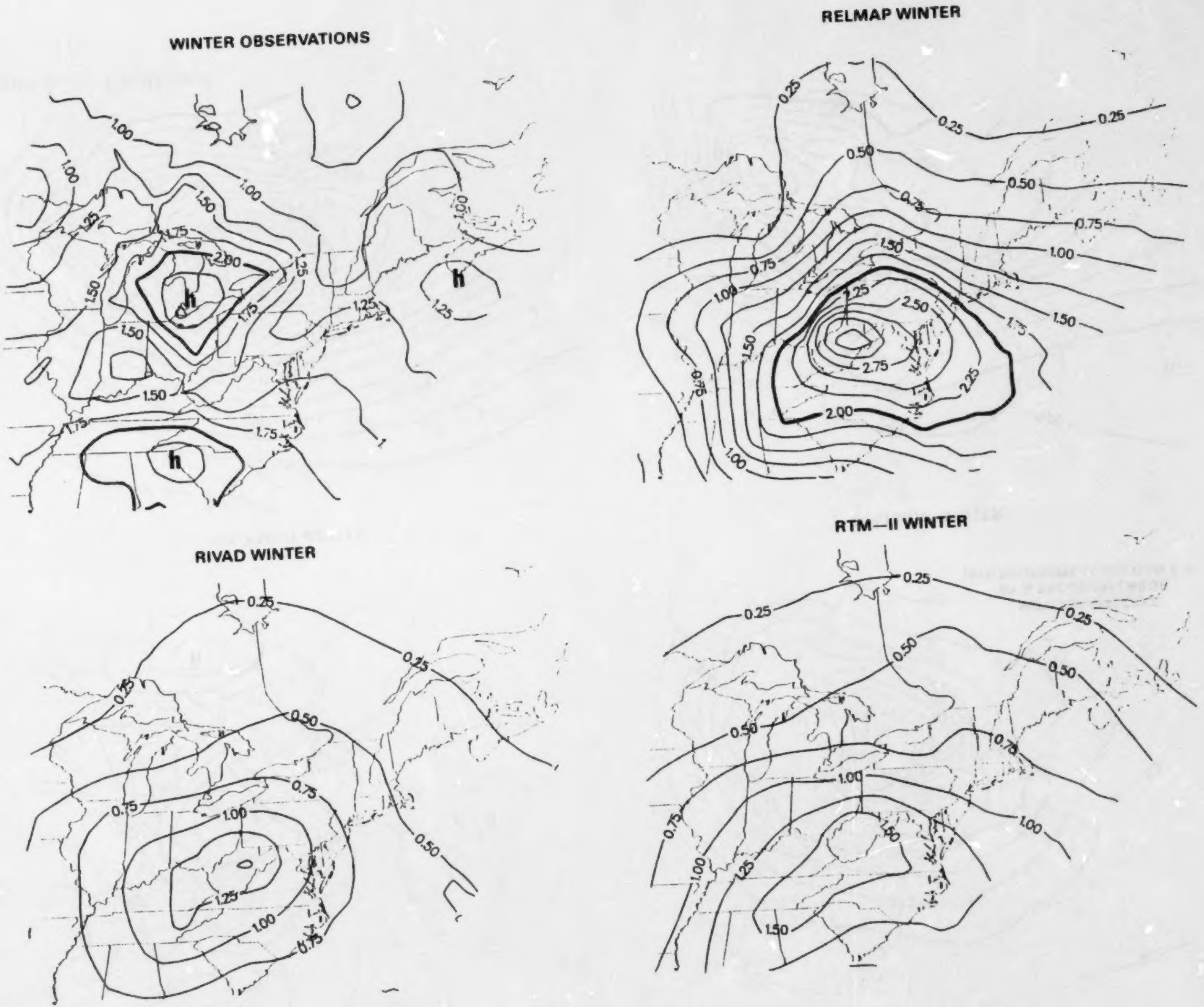

WINTER OBSERVATIONS

AES WINTER

ASTRAP WINTER

MOE WINTER

Figure 6.34 **Winter 1980** kriged analyses of the predictions and observations of sulfur wet deposition (kg S/ha).


WINTER OBSERVATIONS

RELMAP WINTER

RIVAD WINTER

RTM—II WINTER

Figure 6.34 Continued

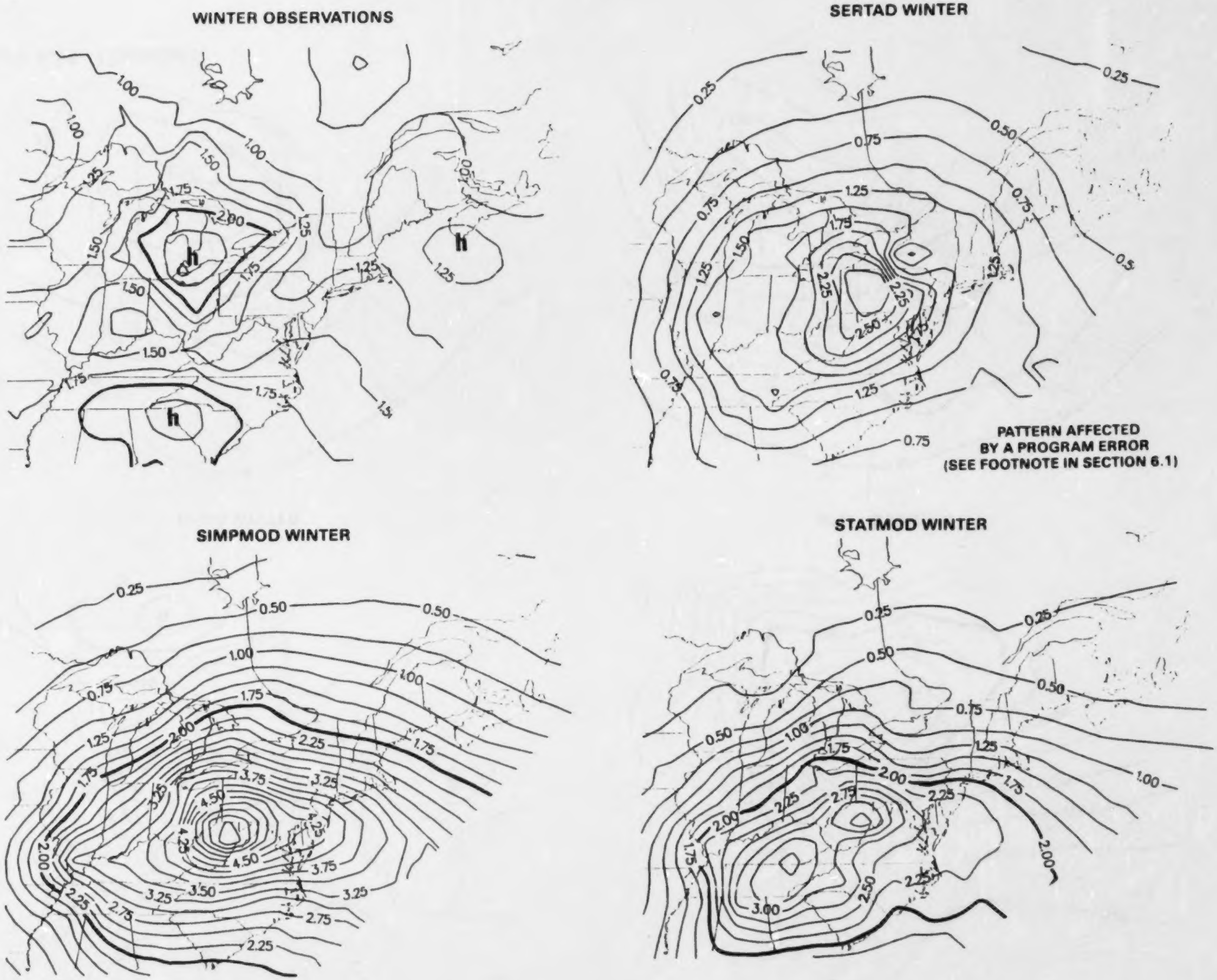

WINTER OBSERVATIONS
SERTAD WINTER
PATTERN AFFECTED BY A PROGRAM ERROR (SEE FOOTNOTE IN SECTION 6.1)
SIMPMOD WINTER
STATMOD WINTER

Figure 6.34 Continued

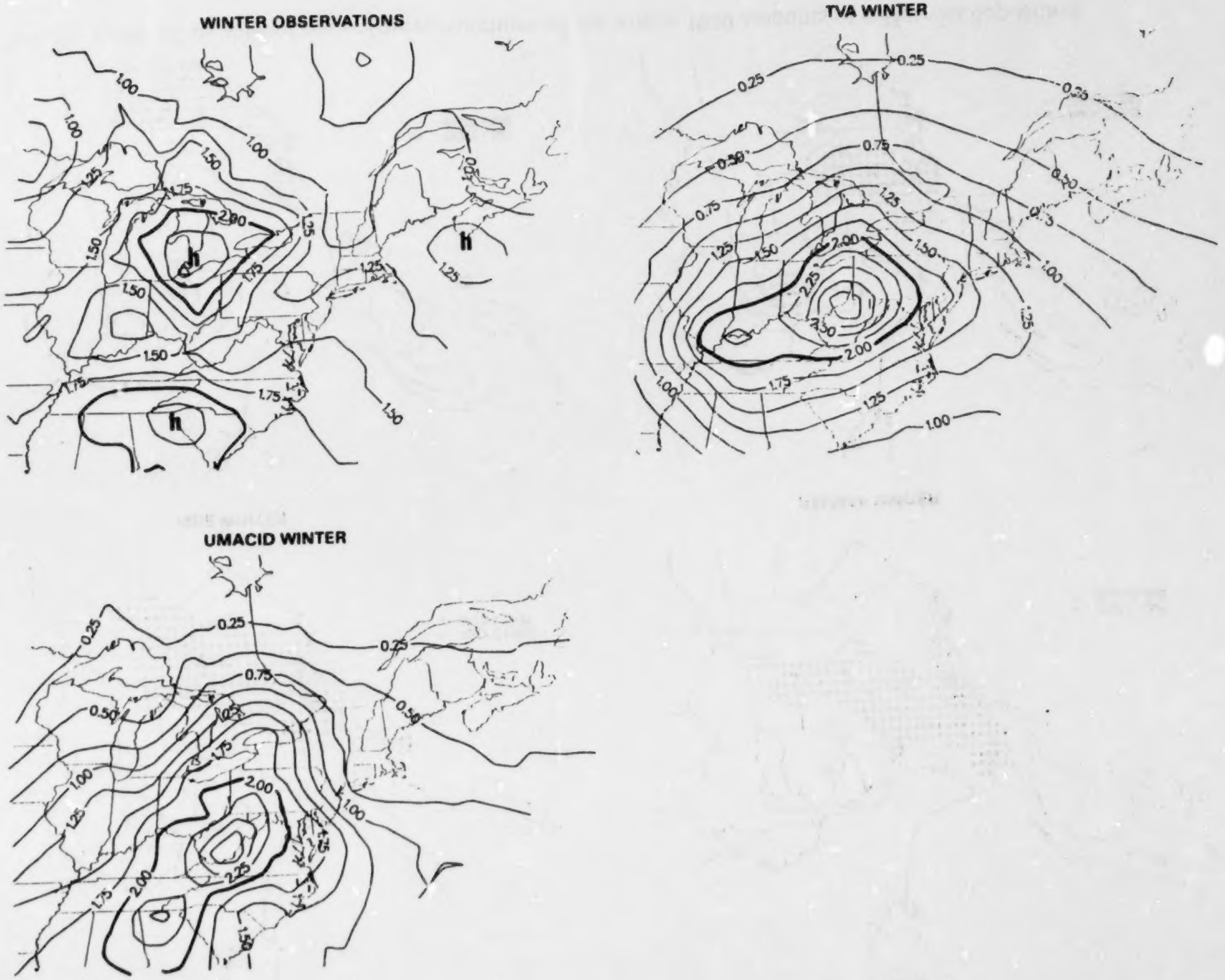


Figure 6.34 Continued

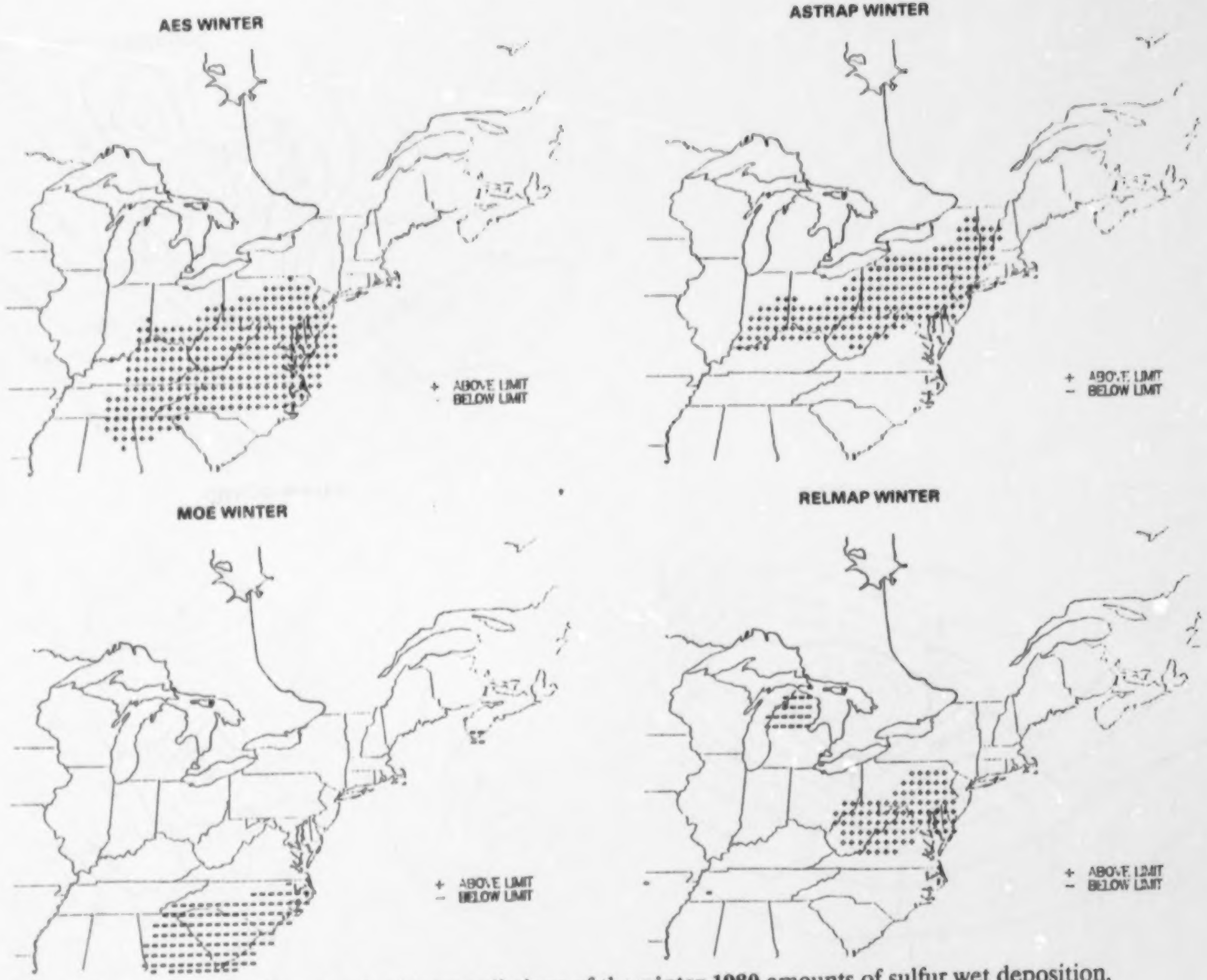


Figure 6.35 Areas of significant under/overpredictions of the **winter 1980** amounts of sulfur wet deposition.

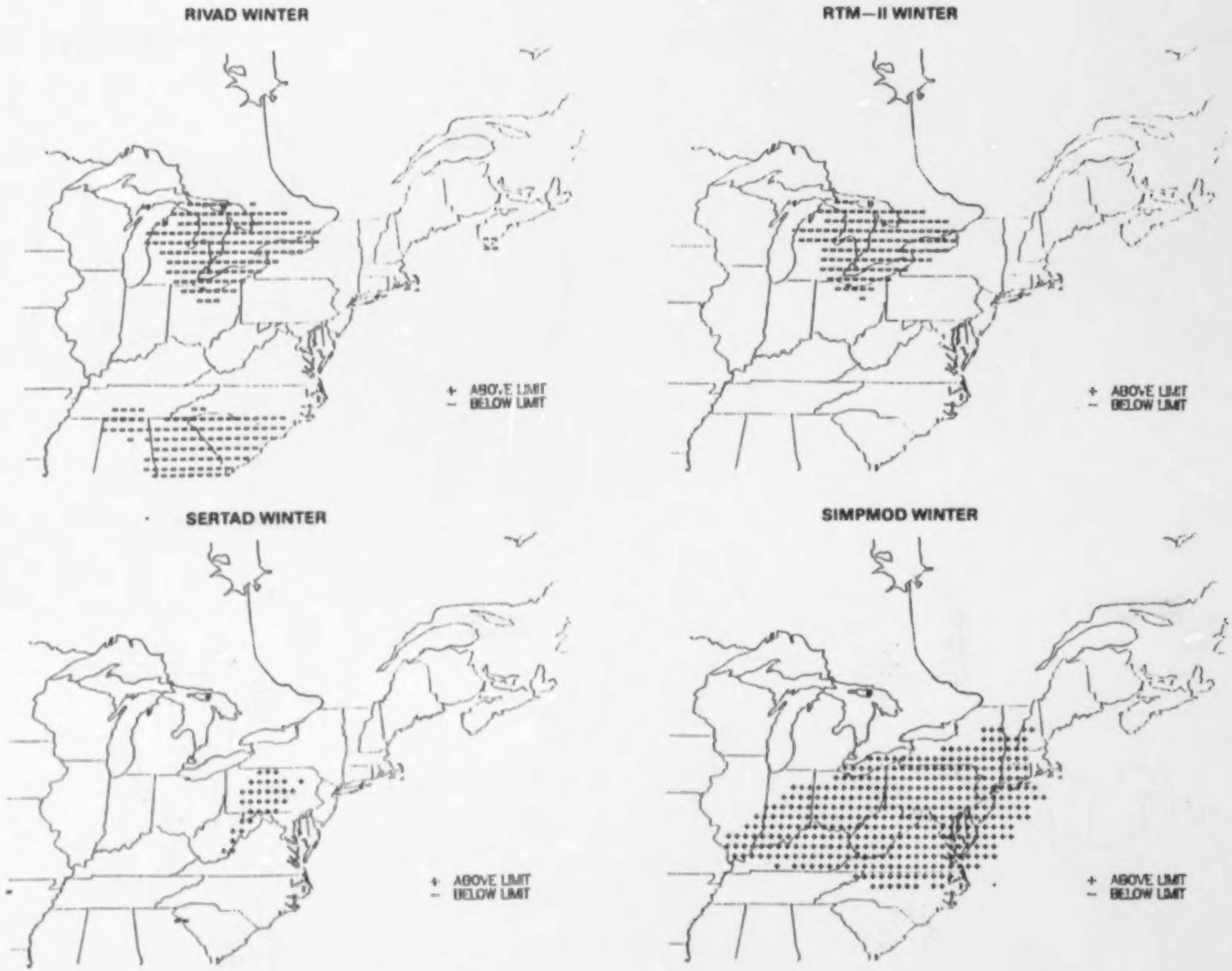


Figure 6.35 Continued

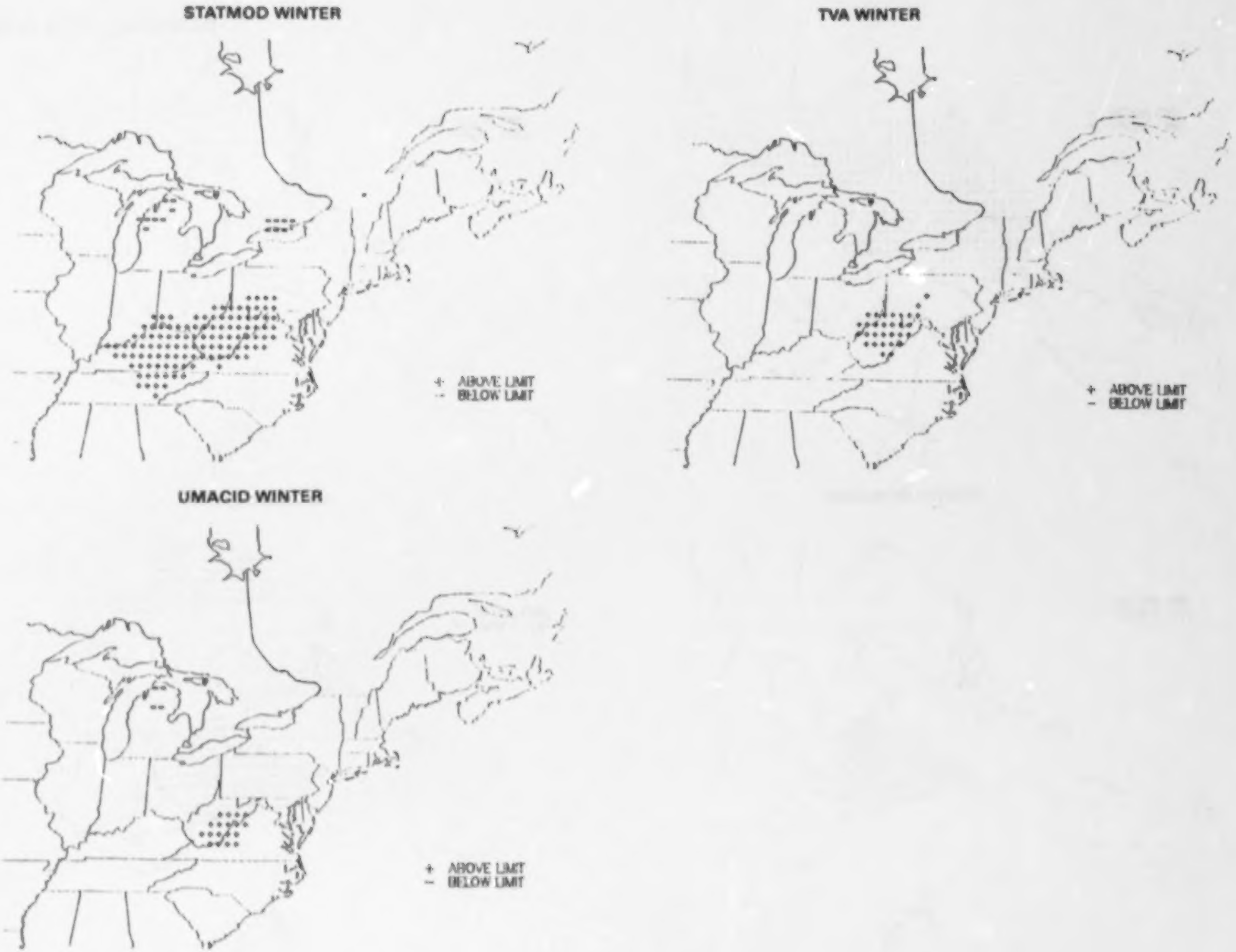


Figure 6.35 Continued

TABLE 6.9 TYPES OF PATTERNS OF SIGNIFICANT DIFFERENCES FOR WINTER 1980.

TYPE	DESCRIPTION	MODELS
I	Significant differences nowhere or limited to small widely scattered areas	SERTAD* TVA UMACID
II	Significant overpredictions across areas to the east and/or south of the Ohio River Valley	AES ASTRAP RELMAP SIMPMOD STATMOD
III	Significant underpredictions across continuous or sporatic areas from the midwestern United States to Nova Scotia	RIVAD RTM
IV	Significant underpredictions across the southern United States	MOE RIVAD

* See footnote in Section 6.1.

6.3.4.4 Model Performance Measures

Pattern Magnitude - The predictions of nine models (all but AES and SIMPMOD) significantly differ from the observations in less than 20% of the composite region (Table 6.10). Further, the predictions of MOE, SERTAD, TVA and UMACID significantly differ in less than 5% of the composite region and not more than 25% of any subregion. Only these four models receive the maximum score of four for pattern magnitude, since the others have significant differences in more than 25% of at least one subregion.

Three of these nine models have predictions which significantly differ from the observations in over 25% of only one subregion and, therefore, receive a score of two. RELMAP significantly overpredicts across half of the southeast subregion. Both RIVAD and RTM significantly underpredict across the north central subregion.

TABLE 6.10 SUMMARY OF MODEL PERFORMANCE MEASURES FOR WINTER 1980.

	AES	AST	MOE	REL	RIV	RTM	SER	SIM	STA	TVA	UMA	OBS
I. Spatial Pattern Replication												
A. Magnitude												
1. Area of Significant Differences (%)*												
a. Northeast	0.0	0.0	0.8-	0.0	0.8-	0.0	0.0	0.0	0.0	0.0	0.0	
b. North Central	0.2+	11.5+	0.0	5.4-	31.2-	27.6-	0.7+	14.7+	3.6-	0.0	0.9-	
c. Southwest	41.2+	29.4+	0.0	5.3+	4.4-	2.6-	0.0	62.3+	33.3+	4.4+	1.8+	
d. Southeast	85.5+	52.6+	5.3-	50.7+	0.0	0.0	19.7+	96.7+	32.7+	13.2+	9.2+	
e. All 4 regions	20.7+	18.2+	0.9-	10.4+	13.8-	11.8-	3.0+	32.5+	11.7+	2.8+	2.1+	
2. Median (kg S/ha)	2.38	2.25	1.15	1.58	0.87	0.99	1.50	2.97	2.02	1.61	1.49	1.34
(p-value)	0.009	0.000	0.097	0.477	0.004	0.009	0.477	0.000	0.235	0.477	0.477	
3. Dispersion (kg S/ha)	2.12	0.75	0.29	0.78	0.13	0.17	0.50	2.01	1.05	0.59	0.58	0.44
(p-value)	0.000	0.064	0.113	0.052	0.000	0.003	0.359	0.000	0.007	0.203	0.212	
4. Bias-High (kg S/ha)	-0.77	-1.06	0.18	-0.08	0.78	0.72	-0.12	-1.87	-0.22	-0.29	-0.02	
(p-value)	0.03	0.00	0.18	0.38	0.00	0.00	0.29	0.00	0.25	0.13	0.46	
5. Bias-Low (kg S/ha)	-0.91	-0.48	0.44	-0.10	0.54	0.29	0.09	-1.08	-0.27	0.06	0.06	
(p-value)	0.00	0.00	0.00	0.29	0.00	0.01	0.28	0.00	0.05	0.35	0.37	
6. Mean Squared Error(kg S/ha)^2	2.85	1.33	0.64	0.88	0.85	0.63	0.62	4.10	0.95	0.65	0.55	
B. Position												
1. Centroids (km)	146	43	82	144	66	112	42	81	151	32	87	
2. Major Axis (degrees)**	+5.3	+8.7	+13.3	+3.7	+10.2	+10.2	+9.6	+12.6	+20.1	+14.4	-0.1	37.5
3. Shift (km)	704	492	245	704	641	737	467	467	660	397	592	
C. Shape												
1. Correlation-No Shift	-0.17	0.03	0.36	-0.16	-0.06	-0.11	-0.18	-0.06	-0.02	0.11	0.08	
2. Correlation-With Shift	0.79	0.68	0.61	0.74	0.73	0.78	0.63	0.58	0.65	0.69	0.74	
II. Maxima												
A. Location-Kriged (km)	238	42	55	238	202	202	197	238	238	101	253	
B. Magnitude-Kriged (%)**	+36	0	0	0	-25	-18	0	+38	0	0	0	
Magnitude-Site (%)**	+59	+7	-19	0	-42	-37	0	+63	+1	0	0	
III. Temporal Pattern Replication												
A. Seasonal Contribution (%)	25.0	19.4	16.2	13.7	18.1	20.5	16.3	32.0	26.7	19.3	22.5	16.2
(Percent Difference)**	+54	+20	0	-15	+12	+27	+1	+98	+65	+19	+39	

Underscored measures are primary performance measures.

* "+" and "-" denote the prevalence of overpredictions and underpredictions, respectively.

** Values listed for each model represent the value of the prediction minus the value of the observation. Percent differences are normalized by observed values.

The remaining four models (AES, ASTRAP, SIMPMOD and STATMOD) score zero points for pattern magnitude. AES and SIMPMOD significantly overpredict across more than 40% of the southwest and southeast subregions. ASTRAP and STATMOD significantly overpredict across more than 25% of both the southwest and southeast subregions.

For winter, the descriptive statistics alone can generally distinguish between those models receiving four and zero points for pattern magnitude. With few exceptions, the four models scoring four points for pattern magnitude (MOE, SERTAD, TVA and UMACID) have the highest p values for the median, dispersion and biases (i.e., the lowest probability that these measures for the predictions are significantly different from those of the observations) and the lowest mean squared errors.

However, on the basis of the descriptive statistics alone, RELMAP, despite its significant overpredictions across half of one subregion, and STATMOD, despite its significant overpredictions across a third of two subregions, would appear to perform as well as those models which significantly under/overpredicted across only 5% or less of the composite region.

For STATMOD, the bias across the high and low deposition regions (based on ISDME site values) is low (Table 6.10), yet at the few sites south of the Ohio River the residuals are much greater (see Table D.7). The overpredictive nature of this model across this large region (Figure 6.34) is masked simply by the sparseness of the monitoring networks.

There are no affects on the pattern magnitude scoring of ASTRAP, since it receives a score of zero for the winter. Likewise, the RELMAP score is not affected, since the variance of its winter predictions is small compared to that of the observations.

Pattern Position - Similar to autumn, the pattern position of the winter predictions generally does not compare favorably with that of the observed. However, two models (ASTRAP and SERTAD) pass the criteria for both the centroid and the orientation of the major axis. Both predict the centroids within 75 km of the observed centroid (Figure 6.36) and orient their major axes within 10 degrees of that of the observations.

Two other models (RIVAD and TVA) pass only the criterion for the centroids. On the other hand, AES, RELMAP and UMACID pass only the orientation criterion. The remaining models (MOE, RTM, SIMPMOD and STATMOD) pass neither criterion for pattern position. The relatively large differences in the orientation of the major axes of the MOE, SIMPMOD and STATMOD predictions and observations reflect the west-to-east ridges of high values depicted in the predicted patterns illustrated in Figure 6.34.

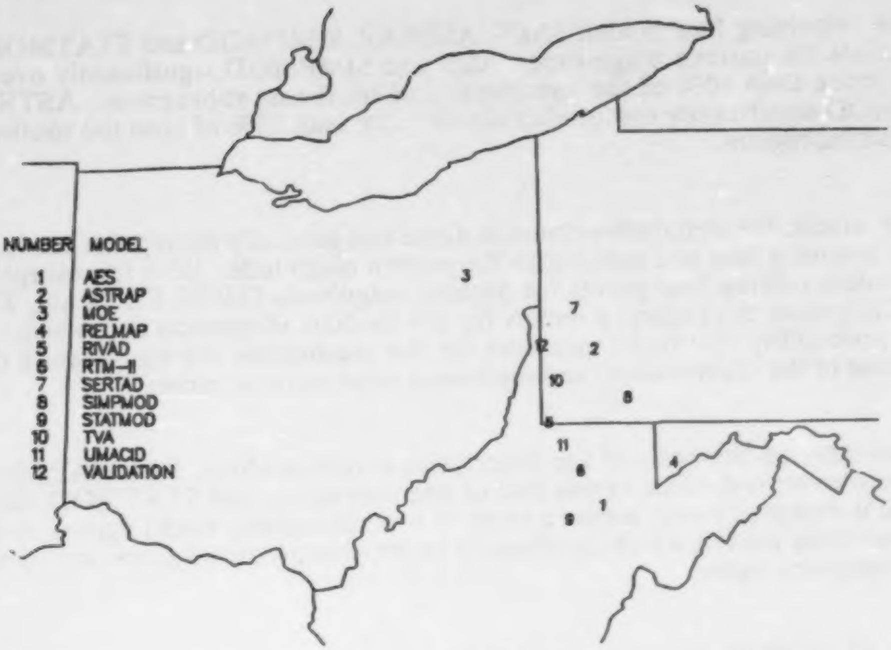


Figure 6.36 Winter 1980 locations of the centroids.

Similar to the other three seasons, as Figure 6.37 illustrates, the difference between the orientation of the major axes of the winter observations and predictions is greatest for STATMOD. Also, similar to the summer and autumn, the orientation of the major axes of two other statistical models (SIMPMOD and TVA) as well as MOE is rotated clockwise 11 to 21 degrees.

Since the orientation of the major axis likely reflects the mean transport vectors during precipitation events, the approach of the statistical models in simulating transport (i.e., long-term wind roses) apparently is not as accurate as the deterministic approach of the other models that use rawinsonde data. These long-term wind roses could represent the frequency distribution of wind velocities during the entire period, but could differ for those periods consisting only of precipitation events. The MOE model is the only deterministic model not using rawinsonde data.

Figure 6.37 also illustrates the seasonal variability of the orientation of the major axes of the observations. It is not known how the orientation varies seasonally for years other than 1980. In both the winter and summer of 1980 the major axis is oriented approximately 40 degrees from North. However, in spring the orientation is 50 degrees while for autumn it is 30 degrees. With the exception of autumn, UMACID duplicates this pattern. AES, RELMAP, RIVAD and RTM mimic general features of the pattern. However, for the other models, the seasonal variations of the orientation of the major axes are confined to a narrow band of 5 degrees.

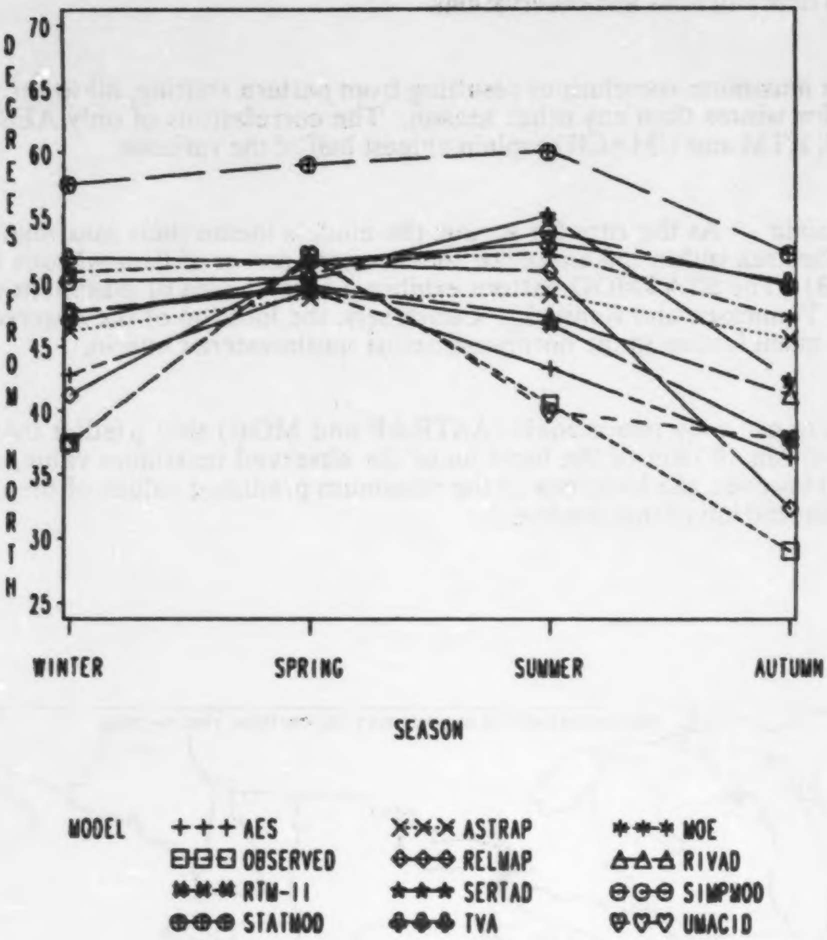


Figure 6.37 Seasonal variation of the orientation of the major axes of the annual 1980 predictions and observations of sulfur wet deposition.

The high values for the performance measure of pattern shift to maximize the correlations provide additional evidence of the poorer winter comparison of predicted to observed pattern positions. Predicted patterns are shifted in excess of 450 km for all but two models (MOE and TVA). One factor explaining the higher winter values is the fact that the effects of pattern rotation are not considered here. As can be seen from Figure 6.34, most predicted patterns appear to be substantially rotated from the observed pattern. Because of its potential to yield inconsistent results, this measure is not used to score the performance of the models. It can, however, yield results that confirm what is apparent from the patterns (e.g., the winter pattern of UMACID).

Pattern Shape - Like autumn, the winter correlations are very low. The MOE correlation again is the highest, although it is less than 0.4. All other correlations are less than 0.2, indicating the poor agreement in the shape and/or position of the patterns of predictions and observations.

The maximum correlations resulting from pattern shifting, all lower than 0.8, are lowest for winter than any other season. The correlations of only AES, RELMAP, RIVAD, RTM and UMACID explain at least half of the variance.

Maxima - As the case for spring, the models locate their maximum predictions across the area within 150 km of the southwestern corner of Pennsylvania (Figures 6.34 and 6.38). The STATMOD pattern exhibits a second area of maximum values across eastern Tennessee and Kentucky. Conversely, the location of the observed maximum value is much further to the northwest, across southwestern Ontario.

There are only two models (ASTRAP and MOE) that predict their maximum values within 100 km of the location of the observed maximum value, the criterion value. However, the locations of the maximum predicted values of the other models are within 260 km of that observed.

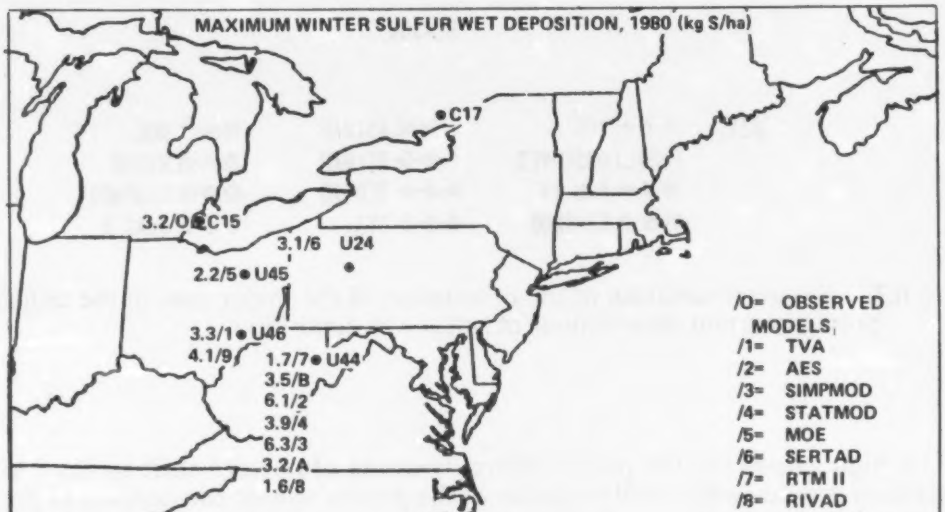


Figure 6.38 **Winter 1980** locations and magnitudes (kg S/ha) of the maximum amount of sulfur wet deposition at the ISDME sites.

The same two models, as well as RELMAP, SERTAD, STATMOD, TVA and UMACID, predict the magnitude of the observed maximum value within the uncertainty range. The magnitudes of the RIVAD and RTM maxima are within 30% of the uncertainty range as determined by the kriging approach. Both approaches indicate that the magnitude of the AES and SIMPMOD maximum predictions is more than 30% greater than the upper limit of the uncertainty range. Therefore, nine models pass the criterion for the magnitude of the winter maximum value.

Temporal Pattern - The range of the relative contributions of winter predictions, 12.5 to 32.0%, is highest for winter. The SIMPMOD winter contribution doubles the 16% observed contribution, while AES and STATMOD winter contributions exceed that of the observed by over 50%. Only the MOE, RELMAP, RIVAD and SERTAD winter contributions are within 15% of that observed and, therefore, only these models pass the criterion for this measure.

6.4 Consistencies of Seasonal and Annual Results

The eight deterministic models of the ISDME compute annual predictions by summing or averaging predictions for the four seasons. The statistical models, which rely on mean durations and frequencies of wet and dry periods, compute seasonal and annual quantities independently. Thus, for the latter models, the aggregation of the seasonal predictions are inconsistent with the annual predictions.

If model predictions are to be used for the evaluation of control strategy scenarios and for the assessment of ecological effects on a seasonal as well as an annual basis, it is essential that the seasonal predictions are consistent with the annual predictions. A simple check of seasonal and annual amounts of predicted wet deposition is performed to identify those models which have annual results inconsistent with the seasonal results. At each of the ISDME evaluation sites, the difference is computed between the sum of the seasonal and the annual sulfur wet deposition predictions. Minor differences can be attributed to round off errors.

All models except SERTAD, STATMOD and TVA (all of which are statistical models) have a mean difference of less than 3% *and* at any site a difference of less than 10%. The mean differences for SERTAD and STATMOD are within 5%, but that for TVA is 11.3%. The site differences for these models are as great as 15 to 25%.

7. Model Clustering

7.1 Introduction

The preceding section discussed the values of the model performance measures for each season of 1980. As mentioned in Section 4.4, the values of six of these model performance measures, referred to as primary performance measures, namely,

1) Area of significant differences (pattern magnitude)

2) Difference in the location of the centroids (pattern position)

3) Difference in the orientation of the major axis (pattern position)

4) Difference in the location of the maximum value (maximum)

5) Difference in the magnitude of the maximum value (maximum)

6) Difference in the seasonal contributions (temporal pattern)

were compared to the criteria values in Table 4.3.

As a means of comparing model performance, seasonal scores were assigned to the models on the basis of the number of these primary model performance measures passing the prescribed criteria (see Table 4.3). With the exception of the first primary measure in the list above, a model received one point for each primary measure passing the criteria. For the measure of the area of significant differences, a model received a score of four points when the more stringent criteria were satisfied or two points when the less stringent criteria were satisfied. The fact that this measure assesses

1) The significance of the differences between the predictions and observations at every half-degree cell and

2) The differences of more than one attribute of the spatial patterns (i.e., it actually assesses, to varying degrees, pattern magnitude, position and shape)

Preceding page blank

justifies the higher weight, since the other primary measures focus on a single attribute of the patterns.

The maximum score a model can receive for the seasonal and overall performance is nine. For the annual performance, the maximum score is only eight, since the primary measure for temporal pattern is not considered in the assessment.

The remainder of this section intercompares the seasonal, overall and annual performances of the models. The subsequent discussions focus on the overall model performance and the degree of consistency of the seasonal performances of each model. In addition, explanations of the seasonal and overall performances are discussed.

7.2 Overall Model Performance

The assessment of overall model performance is based on overall scores calculated by summing weighted seasonal scores. The set of weights, which reflects the seasonality of the observations of sulfur wet deposition, is determined by spatially averaging the relative seasonal contributions of the observations (see Figure 6.1). Accordingly, the set of weights is

0.39	for summer
0.29	for spring
0.16	for autumn
0.16	for winter.

Consequently, those models which perform well for summer and spring, the seasons when the amounts are greatest, will tend to excel in their overall performance.

As mentioned in Section 4.4, the overall performance is a better gauge of model performance and reliability than only the performance of the model in replicating annual patterns of sulfur wet deposition. An assessment of model performance based on its overall performance, as opposed to only its annual performance, at least partially addresses the question, did the model predict the right values for the right reason? In other words, did the model predict the highest seasonal deposition amounts when they were observed? An assessment based on overall model performance, therefore, will address an additional attribute of the model --- the seasonality of its parameterizations. If a model does not mimic the observed seasonality, then its annual predictions, even though they concur with observations, were probably obtained for the wrong reasons.

Table 7.1 presents the seasonal, overall and annual model performance scores for pattern magnitude and position, maximum value, and temporal pattern, as well as the total score. As illustrated by the distribution of overall scores in Figure 7.1, there appears to be three clusters of models with similar scores. Group A models have overall scores greater than 7, Group B models have scores between 5 and 7 and Group C models have scores lower than 5. The following subsections discuss the performance of the models in each cluster.

7.2.1 Group A Models

MOE, SERTAD and TVA each have overall performance scores greater than 7 as a result of their consistently high seasonal scores. Each model scores the maximum points (4) for the pattern magnitude for each season and MOE and TVA scored the maximum number of points (9) for spring (Table 7.1).

Based only on its performance scores, SERTAD appears to be one of the better performing models (see Footnote). However, the following aspects of the model and, in particular, its application in this "blind test" evaluation study can lead one to wonder if its relatively good performance was achieved for the wrong reasons:

1) The contribution of its seasonal "correction" terms to the prediction of sulfur wet deposition was substantial (e.g., 40 to 50% of the magnitude of its predictions across the high deposition region of North America are attributed to the correction terms). These correction terms were proportional to the product of the seasonal/annual precipitation amounts and a constant seasonal/annual concentration of sulfate in precipitation that was approximately three times greater than those observed at remote sites. In addition, the values of these correction terms were computed from data collected during a span of several years, including the ISDME evaluation year.

2) Its high initial conversion of sulfur dioxide to sulfate near the sources (i.e., 20 to 40% of the mass of sulfur dioxide is immediately converted to sulfate) was much greater than the other models.

3) The model ignores the dry deposition of sulfate.

4) Its predictions of seasonal and annual mean air concentrations of sulfate at the four APN sites are greater than those observed by factors of two to ten. In addition, these predictions at the other ISDME sites are at least a factor of two greater than those of the other models.

Footnote: The effects on the performance scores of a SERTAD postprocessor programming error, brought to the authors' attention just prior to the publication of this report have not been quantified (see footnote in Section 6.1)

TABLE 7.1 THE SEASONAL, OVERALL AND ANNUAL MODEL PERFORMANCE SCORES.

		Pattern Magnitude	Pattern Position	Maximum Value	Temporal Pattern	Total
	Maximum	4	2	2	1	9
SUMMER 1980	AES	4	1	1	0	6
	ASTRAP	2	2	1	1	6
	MOE	4	1	1	1	7
	REIMAP	0	0	0	1	1
	RIVAD	0	2	1	1	4
	RTM	0	2	1	1	4
	SERTAD*	4	1	1	1	7
	SIMPMOD	0	0	0	0	0
	STATMOD	0	0	0	0	0
	TVA	4	0	2	1	7
	UMACID	0	2	1	1	4

		Pattern Magnitude	Pattern Position	Maximum Value	Temporal Pattern	Total
	Maximum	4	2	2	1	9
SPRING 1980	AES	4	2	2	0	8
	ASTRAP	4	2	2	1	9
	MOE	4	2	2	1	9
	REIMAP	2	1	1	1	5
	RIVAD	0	2	1	1	4
	RTM	0	2	1	1	4
	SERTAD*	4	2	2	0	8
	SIMPMOD	4	2	2	0	8
	STATMOD	2	1	2	0	5
	TVA	4	2	2	1	9
	UMACID	2	2	2	0	6

		Pattern Magnitude	Pattern Position	Maximum Value	Temporal Pattern	Total
	Maximum	4	2	2	1	9
AUTUMN 1980	AES	2	1	1	0	4
	ASTRAP	2	1	1	0	4
	MOE	4	1	1	1	7
	REIMAP	4	1	1	0	6
	RIVAD	4	1	0	0	5
	RTM	4	1	0	1	6
	SERTAD*	4	1	2	0	7
	SIMPMOD	2	0	1	0	3
	STATMOD	4	0	1	0	5
	TVA	4	0	1	1	6
	UMACID	4	2	1	0	7

		Pattern Magnitude	Pattern Position	Maximum Value	Temporal Pattern	Total
	Maximum	4	2	2	1	9
WINTER 1980	AES	0	1	0	0	1
	ASTRAP	0	2	2	0	4
	MOE	4	0	2	1	7
	REIMAP	2	1	1	1	5
	RIVAD	2	1	1	1	5
	RTM	2	0	1	0	3
	SERTAD*	4	2	1	1	8
	SIMPMOD	0	0	0	0	0
	STATMOD	0	0	1	0	1
	TVA	4	1	1	0	6
	UMACID	4	1	1	0	6

		Pattern Magnitude	Pattern Position	Maximum Value	Temporal Pattern	Total
	Maximum	4.0	2.0	2.0	1.0	9.0
OVERALL 1980	AES	3.0	1.3	1.1	0.0	5.4
	ASTRAP	2.4	1.8	1.4	0.6	6.2
	MOE	4.0	1.1	1.4	1.0	7.5
	REIMAP	1.5	0.7	0.6	0.8	3.6
	RIVAD	1.0	1.7	0.8	0.8	4.3
	RTM	1.0	1.5	0.8	0.8	4.1
	SERTAD*	4.0	1.5	1.4	0.6	7.5
	SIMPMOD	1.5	0.6	0.7	0.0	2.8
	STATMOD	1.2	0.3	0.9	0.0	2.4
	TVA	4.0	0.7	1.7	0.8	7.2
	UMACID	1.9	1.8	1.3	0.4	5.4

		Pattern Magnitude	Pattern Position	Maximum Value	Temporal Pattern	Total
	Maximum	4	2	2	-	8
ANNUAL 1980	AES	2	1	1	-	4
	ASTRAP	2	1	1	-	4
	MOE	4	2	2	-	8
	REIMAP	2	1	0	-	3
	RIVAD	0	2	1	-	3
	RTM	0	2	1	-	3
	SERTAD*	4	2	1	-	7
	SIMPMOD	2	1	1	-	4
	STATMOD	0	0	1	-	1
	TVA	2	1	2	-	5
	UMACID	2	2	1	-	5

Reproduced from
best available copy.

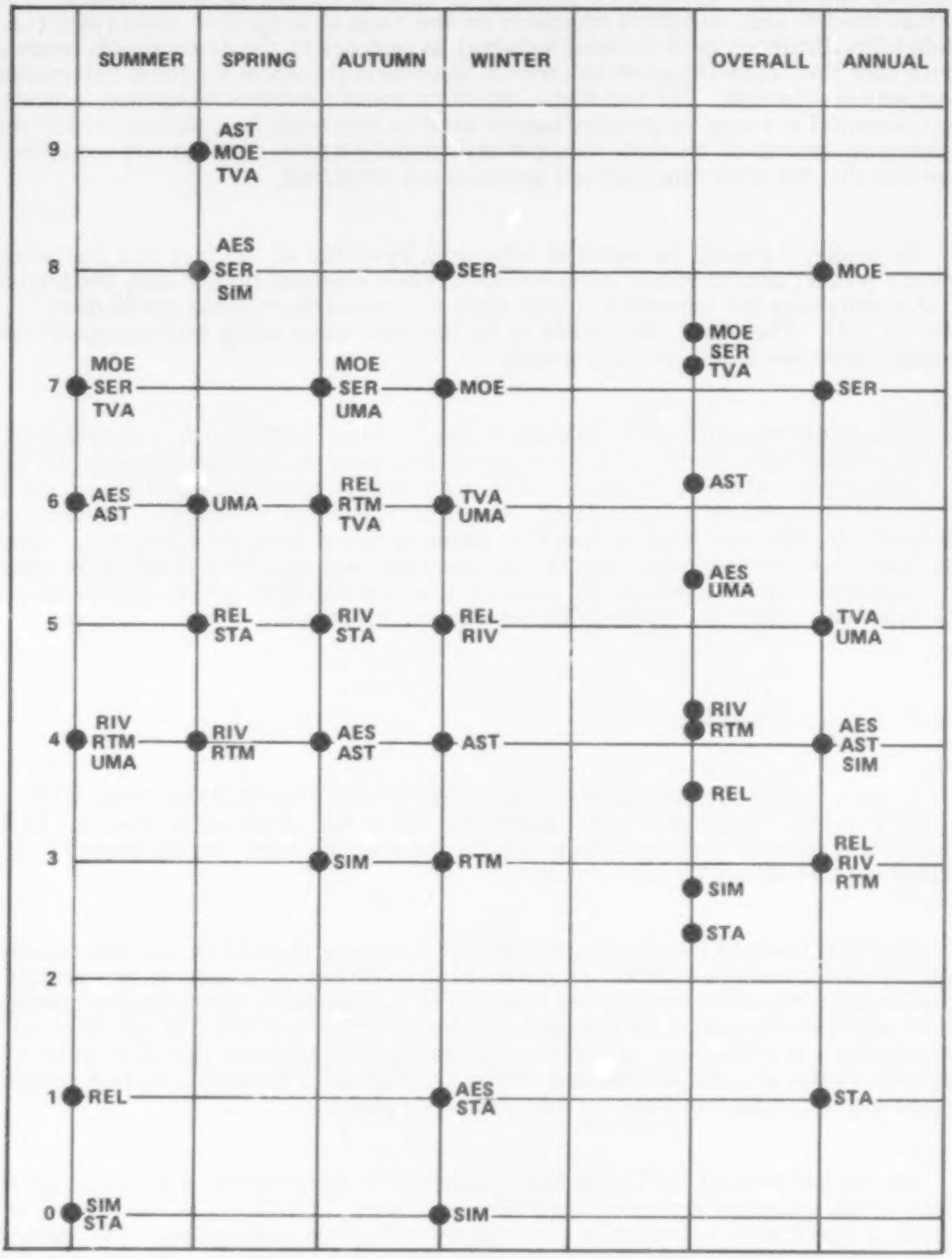


Figure 7.1 The distribution of the seasonal, overall and annual 1980 model performance scores.

Several additional points concerning the Group A models are noteworthy. First, a statistical model (TVA) performs better or as good as the deterministic models. The former models simulate mass transport on the basis of long-term wind roses (i.e., probability distributions of the wind velocity), as opposed to the deterministic models, which base their mass transport on series of short-term trajectory segments determined from actual wind data. The fact that a statistical model performs rather well indicates that seasonal average wind roses can be used to successfully simulate, within the uncertainty bounds of the data, seasonal and annual patterns of sulfur wet deposition, provided that the other processes are appropriately simulated.

Secondly, it should be restated here that, by virtue of the fact that statistical models predict annual values independent of their seasonal predictions, the annual TVA predictions are inconsistent with their accumulated seasonal predictions (see Section 6.4). Therefore, one needs to be cautious when using both seasonal and annual predictions from statistical models.

Finally, MOE, which is the only deterministic model in Group A, is also the only deterministic model in this model evaluation that does not use rawindsonde data to simulate mass transport. Instead, MOE transport is based on geostrophic winds computed from surface pressure fields. Also, it is only one of two models to use a finer resolution for the precipitation data (3-h amounts versus those for 12 or 24 h). Since the rates of the physical and chemical processes are very similar to those of the other deterministic models, perhaps the method in which transport and precipitation are simulated can explain the relatively higher score of this model.

7.2.2 Group B Models

Unlike the Group A models, AES, ASTRAP and UMACID perform well for several seasons, but perform rather poorly for one or two of the other seasons. Their inconsistent seasonal performances precludes these models from earning overall scores as high as those for the Group A models.

Both AES and ASTRAP perform very well for spring (8 and 9 points, respectively) and relatively well for summer (6 points each). However, primarily because they significantly overpredict across more than half of the southeast subregion, both models score only 4 or less points for autumn and winter. The overpredictive nature of AES for autumn and winter can be explained by the lack of seasonal variation of its wet deposition rates of sulfur dioxide and sulfate. As Figure 6.1 illustrates, there is a strong seasonality in the observations that AES does not mimic.

On the other hand, ASTRAP has a tendency to significantly overpredict across much of the southeast subregion regardless of season. In addition to the autumn and winter, it significantly overpredicts across nearly half of the same subregion in summer. Since this subregion is typically downwind of the high emission regions, it seems that this model removes sulfur at a rate significantly higher than the actual rate. However, this model does not appear to significantly underpredict across the areas further downwind, which is what one might expect. ASTRAP does tend to underpredict at distant sites (see Tables D.7-D.10), however, the differences are insignificant when the uncertainties of the observations are considered. The magnitude of these uncertainties relative to the observations are highest where the density of sites is lowest, and for this case, where the model underpredicts.

Unlike AES and ASTRAP, UMACID performs best in autumn, winter and spring (6 or 7 points each). Its summer performance (4 points) is marred by significant underpredictions across 20% of the composite region and particularly across the Great Lake states (nearly a third of the southwest subregion) and Maritime provinces. This model also significantly underpredicts across nearly a third of the southwest subregion in spring. Like AES, the UMACID rates of the wet removal of sulfur dioxide and sulfate do not vary seasonally. Apparently, the prescribed rates are appropriate for several, but not all, seasons.

7.2.3 Group C Models

On the basis of their low scores for overall model performance, the remaining five models (RELMAP, RIVAD, RTM, SIMPMOD and STATMOD) are classified as Group C models. All but RIVAD and RTM perform relatively well for only one season. The two exceptions perform consistently poorly for each season.

RELMAP performs best in autumn when the model receives the maximum score for pattern magnitude. However, the model performs poorly in replicating the pattern position, location of the maximum value and the autumn contribution to the annual total.

The low overall performance score from RELMAP is likely attributable to its high rate of the wet removal of sulfur, as demonstrated by its tendency to significantly overpredict in the southeast subregion. RELMAP significantly overpredicts across 90%, 60% and 50% of the southeast subregion in summer, spring and winter, respectively. In autumn, however, the model significantly overpredicts across less than 10% of the same subregion. The inconsistency in the model performance can be explained in part by the variation in the seasonal precipitation patterns, illustrated by Figures B.3-B.6. Across the subregion where RELMAP tends to overpredict, the autumn precipitation amounts are much lower than those for the other seasons (approximately half those for summer and spring and one-third those for winter). As a consequence, the effects of the high removal rates diminishes for this season.

SIMPMOD performs very well for spring (8 points), but poorly for the other seasons (0 to 3 points). In spring, this model does not significantly under/overpredict anywhere, replicates well the pattern position and predicts well the location and magnitude of the maximum value. However, for summer, this model significantly *under*predicts across 30% of the composite region and across 30 to 40% of each subregion but the southeast. In contrast, for both autumn and winter, it significantly *over*predicts across nearly all of the southeast subregion (nearly 80% and 100%, respectively) and across large portions of the southwest subregion (20% and 60%, respectively).

Since SIMPMOD does not directly use precipitation amounts, the reasons are not apparent for its inconsistent seasonal performances. The fact that, for every season, SIMPMOD (and STATMOD) wet removal rates of sulfur dioxide are an order of magnitude lower, while those for sulfate are factors of two to three lower than those of the other statistical models (SERTAD and TVA) would appear to explain the underpredictions in summer. However, this explanation is not valid, since for spring the model performs well.

The tendency of SIMPMOD to overpredict in autumn and winter could be related to the manner in which the precipitation statistics (i.e., the mean duration, frequency and amount) were determined. If the statistics reflected the annual period, as opposed to individual seasons, the lower precipitation amounts observed across the United States in autumn and winter would not be considered by the model. Consequently, in autumn and winter the model would overpredict across much of the United States, which it does.

The performance of STATMOD is similar to SIMPMOD. The two models are identical with the exception that the former uses seasonal precipitation rates in simulating wet removal. Like SIMPMOD, STATMOD in summer significantly *under*predicts across large portions of the composite region (40%) and every subregion but the southeast. Also as SIMPMOD, STATMOD significantly *over*predicts across much of the United States in autumn and winter.

In contrast to SIMPMOD, however, STATMOD in spring significantly underpredicts across a large portion of one subregion. The reason for this difference is not apparent.

The low overall performance scores of SIMPMOD and STATMOD can also be attributed to the fact that their spatial patterns appear to be shifted, as the other models, but also rotated more than the others. The rotation, which is visible in the figures of Section 6), is also obvious in the orientation of their major axes. For almost every season, the SIMPMOD and STATMOD major axes are oriented more west to east compared to the general southwest-to-northeast orientation of the observed major axes. Another factor in their low scores is that they both fail to mimic the seasonal variation of the observations (Figure 6.1).

Finally, the poor performance of RIVAD and RTM is attributed to the significant underpredictions across extensive areas of the evaluation region, especially for summer and spring. However, for these two seasons, both models receive maximum scores for the performance measures that do not relate to magnitudes: Their pattern positions, locations of the maximum values and the temporal patterns correspond to those of the observations. This suggests that there is a systematic problem with the wet removal rates. Since their predictions of sulfur dry deposition are higher than those of the other models, perhaps these two models are removing too much mass via dry deposition. In addition to their significant underpredictions, the predicted patterns of both models are excessively smooth.

7.3 Annual Model Performance

In contrast to the overall performance, the annual performance is assessed independently of the seasonal performances. Only annual predictions of sulfur wet deposition are compared to annual observations. The maximum annual performance score is 8 versus 9, since the measure for temporal pattern replication does not apply.

The three Group A models, which have the highest overall performance scores, also have the highest annual performance scores. In addition, MOE scores the maximum number of points. MOE and SERTAD, at least for 1980, predict with equal success the seasonal and the annual sulfur wet deposition amounts. However, TVA does not perform as well for the annual period, which suggests that the wind roses and/or statistical input data used for the annual simulation might not be as appropriate as those for the seasonal simulations.

By varying degrees, all but one of the Group B and C models (SIMPMOD) have annual performance scores lower than their overall performance scores. One explanation is the fact that there is an incommensurate maximum number of points (i.e., the maximum number of points for the overall performance score is one greater).

8. SUMMARY AND CONCLUSIONS

8.1 The ISDME Approach

The International Sulfur Deposition Model Evaluation (ISDME) project, jointly conducted by the U.S. Environmental Protection Agency and Atmospheric Environment Service of Environment Canada, assessed the performance of eleven linear-chemistry atmospheric models for the entire year of 1980. This year was selected for the eastern North American evaluation on the basis of the availability of a spatially and seasonally resolved emissions inventory and an adequate precipitation chemistry data base.

The approach of the study was to first compile and distribute to the participating modelers standardized sets of model input data grids consisting of 6-hour 1000- and 850-mb wind velocities, 12-hour precipitation amounts, climatological mixing heights, and vertically stratified seasonal emissions of sulfur dioxide. The use of this standardized data set by all modelers would minimize the differences in model results arising from differences in model input data. Modelers then applied their models in a "blind" test mode and submitted to the evaluators predictions of wet and dry deposition amounts and air concentrations of sulfur dioxide and sulfate at up to 65 sites.

The evaluation data set used to assess the performance of these models consisted of only sulfur wet deposition amounts, since measurements of other parameters were either nonexistent or sparse. These amounts were calculated from screened precipitation chemistry data from five episodic, weekly or monthly networks. With several slight variations, the screening and calculation procedures were those recommended by the Unified Deposition Data Base Committee established under the Canadian Federal-Provincial Research and Monitoring Co-ordinating Committee. The number of sites passing the screening criteria ranged from 32 for the annual to 46 for the spring periods.

The models were evaluated in an operational mode using a new, more rigorous approach, which involved i) the statistical comparison of spatial patterns rather than sets of point values, ii) the consideration of the measurement and interpolation errors of the observations, and iii) the assessment of seasonal rather than only annual performances. In addition, the more conventional distribution statistics recommended by the American Meteorological Society were computed.

The unique ISDME approach of assessing model performance elucidated the behavior of the models in ways that conventional statistical measures alone could not. For example, in addition to concluding that a model tended to under/overpredict by a specific amount, as conventional statistical measures would yield, the ISDME results also identified the regions where the model predictions were significantly greater or lower than the observations.

Spatial patterns of seasonal and annual amounts of sulfur wet deposition were determined by an interpolation technique known as kriging. This technique quantified the uncertainties in the observations which were used in the evaluation process to identify areas where interpolated predictions were statistically significantly different from the interpolated observations.

Of the many model performance measures that were quantified, six were selected to evaluate the models. They were the area of significant differences and the discrepancies between the observed and predicted centroids, major axis orientation, maximum location, maximum magnitude, and seasonal pattern (i.e., the distribution of the annual amount across the seasons). These measures were perceived to be appropriate for ascertaining the differences of major features of the predictions and observations. Model performance criteria were defined for each of these measures and used in a scoring system to summarize, as well as compare in relative terms, the performance of the models. To supplement the evaluation, predictions of dry deposition amounts and air concentrations of each model were intercompared to identify apparent peculiarities.

8.2 Comparison of Predictions and Observations

The seasonal patterns of the observed amounts of sulfur wet deposition exhibited detailed structure with absolute and secondary maxima. Depending on the season, the location of the absolute maxima ranged from northern West Virginia and central Ohio to southern Ontario. The magnitude of these maxima, also seasonally dependent, ranged from 3.2 (winter and autumn) to 7.9 kg S/ha (summer). Although they might not be significant in light of the data uncertainties, several centers of minimum and secondary maximum values appeared in the spatial patterns.

In contrast, the spatial patterns predicted for each season by the ISDME models in general

1) Resembled patterns of concentric ellipses;

2) Exhibited much less detail than the observed patterns (i.e., the predicted patterns were relatively smoother than the observed patterns and generally did not show secondary maxima);

3) Anchored their area of seasonal and annual maximum values within 150 km of the southwestern corner of Pennsylvania, which coincided with the area of high sulfur emissions. (This indicated that the predictions from these linear sulfur deposition models were dominated by the emissions pattern); and

4) **Have major axes that were oriented more west-to-east than those of the observed seasonal patterns, which generally were oriented southwest-to-northeast.**

Half-degree cells where the interpolated predictions significantly differed from the interpolated observations were identified as one means of objectively assessing model performance. Differences were deemed to be significant at approximately the 95% confidence level when the interpolated prediction fell outside the uncertainty limits of the interpolated observation. This assessment indicated that

1) Interpolated predictions from SERTAD and TVA were within the uncertainty range of the interpolated observations virtually everywhere for each season but autumn for SERTAD and summer for TVA;

2) Interpolated predictions from the remaining models differed significantly from the interpolated observations across certain regions for three or four seasons:

   a) RIVAD and RTM, which predicted very smooth patterns, significantly underpredicted across much of the region from the midwestern United States to Nova Scotia;

   b) ASTRAP, regardless of season, significantly overpredicted across portions of the region over and to the east of the Ohio River Valley (the spatial extent, however, varies seasonally);

   c) RELMAP, similar to ASTRAP, significantly overpredicted across portions of the region to the east and/or south of the Ohio River Valley for every season but autumn;

   d) MOE significantly underpredicted across portions of the southern and midwestern United States to Nova Scotia for every season but spring when virtually everywhere its predictions were within the range of uncertainty of the observations;

   e) AES predicted within the uncertainty limits of the observations virtually everywhere in the spring and significantly underpredicted in summer across portions of the midwestern United States and New Brunswick. The model significantly overpredicted in the autumn across the region from Virginia to New Hampshire and in the winter across the region from northern Alabama to Pennsylvania;

   f) UMACID predicted within the uncertainty limits of the observations virtually everywhere in the autumn and winter. The model significantly underpredicted across a portion of the midwestern United States in the spring and summer. In addition, summer values were significantly underpredicted across the Maritime provinces; and

g) SIMPMOD and STATMOD, which did not replicate the seasonal temporal patterns of sulfur wet deposition,

- o Predicted within the uncertainty range of the observations for one season (spring for SIMPMOD and autumn for STATMOD);
- o Significantly underpredicted across portions of the area from the midwestern United States to Nova Scotia for some seasons (summer for each as well as spring for STATMOD);
- o Significantly overpredicted across portions of the region east of the Ohio River Valley for the other seasons (winter for each as well as autumn for SIMPMOD).

Since at any point the uncertainty range of the interpolated observations is an inverse function of the density of the monitoring networks within the neighborhood of the point, the uncertainty range would diminish as the density of the networks increases. Consequently, for years subsequent to 1980, when more monitoring sites were operating, the uncertainty range is expected to diminish.

8.3 Model Performance Scores

As a means of assessing and intercomparing model performance, criteria were established for six of the ISDME model performance measures (area of significant differences and discrepancies between predicted and observed centroids, major axis orientation, maximum location, maximum magnitude, and temporal pattern). The criteria values of these measures approximated the median value of the distribution of the seasonal values for all models.

For each season, model performance scores were assigned to the models on the basis of the number of these performance measures passing the prescribed criteria. The maximum score was nine (i.e., four points for the measure of the area of significant differences and one point for each of the five remaining measures). Overall performance scores based only on the seasonal performances were determined from a weighted sum of seasonal performance scores. The seasonal weights were determined from the mean relative contributions of each season to the observed annual amount of sulfur wet deposition (i.e., approximately 0.40 for summer, 0.30 for spring, and 0.15 for both winter and autumn).

The distribution of overall performance scores identified the following three groups, or clusters, of models:

1) Group A models, namely MOE (deterministic) and SERTAD and TVA (both statistical), having overall performance scores exceeding 7, performed relatively well for each season, especially for spring.

   MOE performed somewhat better than the other deterministic models. Since the transformation and removal rates were similar to those of the other deterministic models, a possible explanation for the differences in the predictions could relate to the different manner in which transport and precipitation were simulated by this model.

   Although SERTAD predictions of sulfur wet deposition were within the uncertainty limits of the observations, several unique features of the model deserve comment. First, the values of the "correction" (background) terms, used in spring, summer, autumn and annual simulations, were proportional to the product of the seasonal/annual precipitation amounts and constant seasonal/annual background concentrations of sulfate in precipitation that were approximately three times greater than those observed at remote sites. These terms contributed 40 to 60% of the model's predicted sulfur wet deposition, which was considerably greater than that of the only other model that considered a background term. Secondly, SERTAD sulfate air concentrations were always greater than the predictions of other models and greater by a factor of two to ten than the observations at the four APN sites. These effects were attributed to the high initial conversion of sulfur dioxide to sulfate near the source (20 to 40%) and the lack of a dry deposition sink term for sulfate.

2) Group B models, AES, ASTRAP and UMACID, having overall performance scores between 5 and 7, performed relatively well for two or three seasons. The lack of a seasonal variation in the rates of sulfur wet removal for AES and UMACID were likely factors in their inconsistent performance. The ASTRAP rates of sulfur wet removal appeared to be too high, since its overpredictions of sulfur wet deposition occurred across and to the east of the major source region along the Ohio River Valley.

3) Group C models, RELMAP, RIVAD, RTM, SIMPMOD and STATMOD, having overall performance scores of less than 5 points, performed relatively well for only one or two seasons. Based on its patterns of significant overpredictions, the RELMAP rates of sulfur wet removal appear to be too high for at least summer and spring. Its predictions of sulfur dioxide concentration also were too high. This overprediction was caused by the inadvertent injection of sulfur dioxide emissions into the surface layer at night.

> Although both RIVAD and RTM tended to significantly underpredict amounts of sulfur wet deposition for each season, the models performed relatively well for the measures not pertaining to the magnitude of the predictions. This could be indicative of a systematic error in the models. As evidenced by the fact that their predictions of the air concentrations were lower and sulfur dry deposition were higher than those of the other models, the RIVAD and RTM sulfur budgets were quite different.
>
> Finally, both SIMPMOD and STATMOD scored few if any points for all seasons but spring for SIMPMOD and both spring and autumn for STATMOD. The temporal pattern of the predictions of sulfur wet deposition, which deviated more from that of the observations than the other models, explained the extensive, significant underpredictions for some seasons and the extensive, significant overpredictions for some other seasons. This was indicative of a problem in the seasonal variation of their sulfur wet removal.

There were several aspects of the predictions that were common to each of the statistical models (namely, SERTAD, SIMPMOD, STATMOD and TVA), but were dissimilar to the predictions of the deterministic models. First, the orientation of the major axis of the spatial patterns of sulfur wet deposition for the statistical models tended to be further clockwise from the major axis of the observed spatial patterns than the deterministic models. These differences, greatest for STATMOD, were as much as +20 degrees for every season but spring. Since the orientation of the major axis likely reflects the mean transport vectors during precipitation events, the approach of the statistical models in simulating transport (i.e., via long-term wind roses) apparently yielded different results than the approaches of the other models. These long-term wind roses could represent the frequency distribution of wind velocities during the entire period, but could differ for those periods consisting only of precipitation events.

Another difference between the predictions of the two types of models was apparent in the comparison of the seasonal air concentrations of sulfate at the four APN sites. The predictions of the statistical models generally were higher than those predicted by the deterministic models and also higher than those observed. The reason for this was not clear. No systematic differences were apparent for the predictions of seasonal air concentrations of sulfur dioxide.

Finally, the statistical models, which rely on mean durations and frequencies of wet and dry periods, computed seasonal and annual values independently. Thus, the aggregation of the seasonal predictions were inconsistent with the annual predictions. The prevalent opinion is that the statistical models are somewhat constrained to annual applications because of the limitations of the statistics to long periods of time. However, as this model evaluation indicates, some statistical models predict rather well for seasonal periods.

The scope of a comprehensive model evaluation should include more than an assessment of the performance of a model. Ideally, one needs to ascertain whether those models replicating the spatial patterns did so for the "right reasons". That is, was the pattern replication a consequence of the correct simulation of the effects of all significant processes or can it be attributed to factors which didnot simulate these effects, yet produce patterns which fortuitously replicate the observed patterns? Considering these models as regulatory tools, the issue here was model reliability: If a model is capable of replicating the observed patterns for one year, can it be expected to do so for another period when the meteorological and emissions scenarios differ? Since appropriate air concentration and dry deposition measurements were not available, the ISDME could not comprehensively assess model reliability.

8.4 Concluding Remarks

Model development is an evolutionary process involving iterations of model evaluations and subsequent model refinements. That is, the results of a model evaluation can reveal both the strengths and weaknesses of a specific version of a model and can assist the modeler in improving the model. The ISDME study represents one step of this continuous model development process. By assessing the performance of the eleven sulfur deposition models, the results of this study enable the modelers to advance to the next steps of model refinement and re-evaluation. The results also provide a benchmark with which the performance of the refined models can be compared to quantify the effects of the model improvements.

These evaluation results are based on a single application of a specific version of each of the eleven ISDME models. The performances can be affected by model application errors and coding errors that were previously undetected. Therefore, on the basis of a single model evaluation study, one should not dismiss any one of these models as being unworthy of further consideration. However, the sources of the errors must be identified and eliminated and the credibility of the model must be demonstrated before the model results can be used with confidence.

9. RECOMMENDATIONS FOR FUTURE RESEARCH

Model Input Data

1) Improve the accuracy of the temporal and spatial characteristics of the precipitation data for scales commensurate with those of the models.
2) Estimate the uncertainties and errors in the emission inventories and meteorological input data.

Model Refinements and Applications

1) Explore the reasons why the models performed as they did for 1980.
2) Use the 1980 ISDME data set to improve the model performance.
3) For 1980, develop and intercompare source-receptor relationships.
4) Assess the effects of the uncertainties and errors in the model input data on the predictions, the model performance and the source-receptor relationships.

Model Evaluation Data

1) Enhance the rigor of future evaluations by simultaneously measuring or empirically estimating all predicted variables (e.g., air concentrations, dry depositions and wet depositions of each modelled species) at the same sites in a optimally configured network.
2) Adopt the Unified Deposition Data Base Committee methods for screening precipitation chemistry measurements and computing wet deposition amounts.

Model Evaluations

1) Focus on the replication of spatial patterns in evaluating models. It is necessary to compare predictions and observations on commensurate spatial scales.

2) Quantify uncertainties of the patterns of observations arising from interpolation and measurement errors and account for them in the evaluation procedure.

3) Evaluate long-term models on seasonal rather than only annual scales to increase the rigor of the evaluations and the detail of the results.

4) Evaluate the latest versions of the ISDME models using the 1988 and 1989 monitoring data to be collected during the Regional Model Field Study.

5) Continue to develop methods of quantifying differences in the spatial patterns of observations and predictions, namely, rotation and translation.

Interpretation of Observed Data

1) Accompany spatial patterns of observations with spatial patterns of the relative uncertainties; different sets of data for the same period can produce markedly different patterns.

REFERENCES

Barnes, M.G. (1980) The use of kriging in estimating the spatial distribution of radionuclides and other spatial phenomena,TRANSTAT (Statistics for Environmental Studies) #13, Battelle Memorial Institute, Richland, WA, 22 pp.

Barrie, L.A. and Hales, J.M. (1984) The spatial distribution of precipitation acidity and major ion wet deposition in North America during 1980, *Tellus*, *36B*: 333-355.

Barrie. L.A.; Wiebe, A.; Anluf, K. (1984) Results of the Canadian Air and Precipitation Monitoring Network (APN), Technical Report, AQRT-84-005-T, Atmospheric Environment Service, Downsview, Ontario, Canada.

Bhumralkar, C.M.; Mancuso, R.L.; Wolf, D.E.; Nitz, K.C.; Johnson, W.B. (1981) ENAMAP-1 long-term SO2 and sulfate pollution model: further applications to eastern North America, Final Report, EPA-600/7-81-102, U.S. Environmental Protection Agency, Research Triangle Park, NC, 128 pp.

Bilonick, R. (1983) Risk qualified maps of hydrogen ion concentration for the New York State area for 1966-1978, *Atmos. Environ.*, *17*: 2513-2524.

Briggs, G.A. (1969) Plume rise, USAFC Critical Review Series, TID-26075, NTIS, Springfield, VA.

Briggs, G.A. (1971) Some recent analyses of plume rise observa-tions, *Proceedings of the Second International Clean Air Congress*, H.M. Englund and W.T. Berry, Editors, Academic Press, New York, NY, pp. 1029-1032.

Briggs, G.A. (1972) Discussion of chimney plumes in neutral and stable conditions, *Atmos. Environ.*, *6*: 507-510.

Clark, I. (1979) *Practical Geostatistics*, Applied Science Publishers, London, England, 129 pp.

Clark, T.L.; Coventry, D.H.; Eder, B.K.; Benkovitz, C.M.; Evans, V.A.; Volder, E.C.; and Olson, M.P. (1985) International Sulfur Deposition Model Evaluation - 1980 standard model input data set, Paper 85-5.3, 78th Annual Meeting of the Air Pollution Control Association, Detroit, MI, June 16-21.

Clark, T.L.; Dennis R.L.,; Palka, D.; Olson, M.P.; and Voldner, E.C. (1986) Effects of gridded precipitation bias on wet deposition predictions. Paper 86-29.5, 79th Annual Meeting of the Air Pollution Control Association, Minneapolis, MN, June 22-27.

Concord Scientific Corporation (1985) Development of SERTAD: a modeling system for statistical estimates of regional transport and acidic deposition, Part II, Final Report for the Canadian Electrical Association.

Conover, W.J. (1971) *Practical Nonparametric Statistics*, John Wiley and Sons, New York, NY, 462 pp.

Davies, D. (1967) Three-layer numerical forecasts of precipitation amounts. Canadian Meteorological Memoirs, No. 25, Canadian Department of Transport.

Delhomme, J.P. (1978) Kriging in the hydrosciences, *Adv. Water Res.*, *1*: 251-266.

Dennis, R.L. and Seilkop, S.K. (1986) The use of spatial patterns and their uncertainty estimates in the model evaluation process. Proceedings of the Fifth Joint Conference on Applications of Air Pollution Meteorology, November 18-21, Chapel Hill, NC, 41-44.

De Pina, R.G.; Pena, J.A.; and Bowersox, V.C. (1980) Precipitation Collectors Intercomparison Study, Department of Meteorology, Pennsylvania State University, State College, PA.

Eliassen, A. (1978) The OECD study of long range transport of air pollutants: long range transport modeling. *Atmos. Environ.*, *12*: 479-487.

Eynon, B. and Switzer, P. (1983) The variability of rainfall activity, *Can. J. Stat.*,*11*: 11-24.

Finkelstein, P.L. (1983) The spatial analysis of acid precipitation data, *J. of Climate and Appl. Meteor.*, *23*: 52-62.

Fox, D.G. (1981) Judging air quality model performance - review of the Woods Hole Workshop. *Bull. Amer. Meteor. Soc.*, *62*: 599-602.

Holzworth, G.C. (1967) Mixing depths, wind speeds and air pollution potential for selected locations in the United States, *J. of Appl. Meteor.*, *6*: 1039-1044.

Holzworth, G.C. (1972) Mixing heights, wind speeds, and potential for urban air pollution throughout the contiguous United States, Final Report, AP-101, U.S. Environmental Protection Agency, Research Triangle Park, NC, 118 pp.

Huijbregts, C.J. (1970) Le variogramme des residus, Int. Rep., CMM/N-200, Ecole des Mines, Fountainebleau, France, 22 pp.

Journel, A.G. and Huijbregts, C.J. (1978) *Mining Geostatistics*, Academic Press, New York, NY, 600 pp.

Klug, W. and Lupkes, C. (1985) A comparison between long term interregional air pollution models, NR. 104-025-65, Technische Hochschule, Darmstadt, Federal Republic of Germany.

Lawson, C.L. (1972) Generation of a triangular grid with application to contour plotting, Technical Memorandum No. 299, California Institute of Technology, Pasadena, CA, 24 pp.

Lehmhaus, J.; Saltbones, J.; and Eliassen, A. (1986a) A modified sulphur budget for Europe for 1980. EMEP/MSC-W Report 1/86, Norwegian Meteorological Institute, 19 pp.

Lehmhaus, J.; Saltbones, J.; and Eliassen, A. (1986b) Calculated and observed data for 1980 compared at EMEP measurement stations, EMEP/MSC-W Report 1/86 (Annex), Norwegian Meteoroligical Institute, 103 pp.

Martin, D.O. and Tikvart, J.A. (1968) A general atmospheric diffusion model for estimating the effects of air quality of one or more sources, Paper No. 68-148, 61st Annual Meeting of the Air Pollution Control Association, St. Paul, MN.

McNaughton, D.J. (1980) Initial comparison of SURE/MAP3S sulfur oxide observations with long-term regional model predictions, *Atmos. Environ., 14*: 55-63.

McNaughton, D.J. (1984) Use of currently available precipitation data for regional air quality modeling, *J. Air Poll. Control Assoc., 31*: 250-252.

MOI (1982a) U.S./Canadian Memorandum of Intent on Transboundary Air Pollution, Atmospheric Science and Analysis Work Group 2 Phase III Final Report.

MOI (1982b) U.S./Canadian Memorandum of Intent on Transboundary Air Pollution, Emissions, Costs and Engineering Assessment Work Group 3-B Phase III Final Report.

Munger, J.W. and Eisenreich, S.J. (1983) Continental-scale variations in precipitation chemistry, *Environ. Sci.Technol., 17*: 32A-42A.

National Data Use and Access Laboratories (1971) Technical Document No. GT-1, 20 pp.

Norwegian Meteorological Institute (1984) Report of the Meteorological Synthesizing Centre, West (MSC-W) for 1984", Final Report, EMEP/MSC-W Report 1/84, Oslo, Norway.

Pasquill, F. and Smith, F.B. (1983) *Atmospheric Diffusion: Study of the Dispersion of Windborne Material from Industrial and Other Sources*, Third Edition, Ellis Horwood Limited, Chichester, England, 437 pp.

Perhac, R.M. (1978) Sulfate Regional Experiment in northeastern United States: the 'SURE' program, *Atmos. Environ., 12*: 641-647.

Portelli, R. (1977) Mixing heights, windspeeds, and ventilation coefficients for Canada, Climatological Study No. 31, Atmospheric Environment Service, Downsview, Ontario.

Ruff, R.E.; Nitz, K.C.; Ludwig, F.L.; Bhumralkar, C.M.; Shannon, J.D.; Sheih, C.M.; Lee, I.Y.; Kumar, R.; McNaughton, D.J. (1984) Assessment and evaluation of three regional air quality models, Paper No. 84-46.1, 77th Annual Meeting of the Air Pol-lution Control Association, San Francisco, CA.

Ruff, R.E.; Nitz, K.C.; Ludwig, F.L.; Bhumralkar, C.M.; Shannon, J.D.; Sheih, C.M.; Lee, I.Y.; Kumar, R.; McNaughton, D.J. (1985) Evaluation of three regional air quality models, *Atmos. Environ., 19*: 1103-1115.

Rutherford, I.D. (1977) An operational three dimensional multi- variate statistical objective analysis scheme, Issue No. 1, Notes Scientifiques et Techniques, RPN, Atmospheric Environment Service, Dorval, Quebec.

Sabourin, R. (1975) Application of two methods for the interpolation of the underlying variogram, Proceedings of NATO ASI Geostat. 75, Rome, Italy, Reidel Publishing Corp, Dordrecht, Netherlands.

Schiermeier, F.A. and Misra, P.K. (1983) Evaluation of eight linear regional-scale sulfur models by the Regional Modeling Subgroup of the United States/Canadian Memorandum of Intent Work Group 2. Tranactions of the Air Pollution Control Asso-ciation Specialty Conference on the Meteorology of Acid Deposition, pp. 330-345.

Shannon, J.D.; Kleinman, L.K.; Benkovitz, C.M. (1982) Intercomparison of MAP3S models of long-range transport and deposition, Proceedings of the Third Joint Conference on Applications of Air Pollution Meteorology, January 12-15, San Antonio, TX, 29-32.

Shapiro, S.S. and Wilk, M.B. (1965) An analysis of variance tests for nonlinearity (complete samples), *Biometrika, 52*: 591-611.

Sirois, A. and Voldner, E.C. Estimation of dry deposition velocities in North America, Part II: the meteorological data banks (in press).

Snedecor, G.W. and Cochran, W.G. (1967) *Statistical Methods*, Iowa State University Press, Ames, IA, 593 pp.

Stensland, G.J. and Bowersox, V.C. (1984) Quality assurance in acid precipitation through the use of ion balance calculations, Transactions of APCA International Specialty Conference on Assurance in Air Pollution Measurements, pp. 425-443.

Stewart, D.A.; Morris, R.E.; Liu, M.K. (1983a) Evaluation of long-term regional transport models of interest to the National Park Service, Final Report SYSAPP 83/215, Systems Application Incorporated, San Rafael, CA.

Stewart, D.A.; Morris, R.E.; Liu, M.K. (1983b) Evaluation of an episodic regional transport model for a multi-day sulfate episode, *Atmos. Environ., 17*: 1225-1252.

Thorp, J.M., Scott, B.C. (1982) Preliminary calculations of average storm duration and seasonal precipitation rates for the northeast sector of the United States, *Atmos. Environ., 16*: 1763-1774.

Topol, L.E. and Schwall, R. (1984) Performance of precipitation chemistry measurement systems, Transactions of the Air Pollution Control Association International Specialty Conference on Quality Assurance in Air Pollution Measurements, T.R. Johnson and S.J. Penkala, Editors, pp. 444-454.

Unified Deposition Data Base Committee (1985) A unified wet deposition data base for eastern North America: data screening, calculation procedures, and results for sulphates and nitrates, (1980), Final Report, Prepared for the Canadian Federal-Provincial Research and Monitoring Co-Ordinating Committee, 79 pp.

U.S. Environmental Protection Agency (1987) Rocky Mountain acid deposition model assessment: review of existing mesoscale models for use in complex terrain, Final Report, EPA/600/3-87/013, Research Triangle Park, NC.

Vet, R.J.; Chan, W.H.; and Lusis, M.A. (1981) An intercomparison study of three precipitation sampling networks in Ontario - APIOS, CANSAP and GLPN, Ontario Ministry of the Environment, Report ARB-002-81-ARSP.

Vet, R.J. and Onloch, S.G. (1981) Detailed analyses of the Pennsylvania State University intercomparison of precipitation collectors data, Concord Scientific Corporation, Final Report #52-80, Toronto, Ontario.

Vet, R.J. and Reid, N.W. (1982) A comprehensive evaluation and integration of selected wet deposition data from Canada and the U.S.A., Report No. 110-84, Concord Scientific Corporation, Downsview, Ontario.

Voldner, E.C.; Olson, M.P.; Oikawa, K.; and Loiselle, M. (1981) Comparison between measured and computed concentrations of sulphur compounds in eastern North America, *J. Geophy. Res., 86*, No. C6, 5339-5346.

Voldner, E.C. and Shannon, J.D. (1983) Evaluation of predicted wet deposition fields of sulfur in eastern Canada, Transactions of the Air Pollution Control Association Specialty Conference on the Meteorology of Acid Deposition, October 16-19, Hartford, CT, pp. 387-399.

Voldner, E.C.; Sirois, A.; Clark, T.L. (1984) Data screening and calculation procedures for the North American precipitation chemistry data to be used in the International Sulfur Deposition Model Evaluation, Proceedings of the Air Pollution Control Association/American Society for Quality Control Specialty Conference on Quality Assurance in Air Pollution Measurements, October 14-18, Boulder, CO.

Watson, C.R. and Olsen, A.R. (1984) Acid Deposition System (ADS) for statistical reporting: system design and user's code manual, Final Report, EPA-600/8-84-023, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Weber, A.H. and Kurzeja, R.J. (1985) Proceedings of the DOE/AMS Air Pollution Model Evaluation Work Shop, volume 2, statistical analysis of results, Savannah River Laboratory, E.I. du Pont de Nemours and Co., DP-1701-2, Aiken, SC.

APPENDIX A

SUMMARY OF PREVIOUS EVALUATIONS
OF SULFUR DEPOSITION MODELS

TABLE A.1 SUMMARY OF PREVIOUS EVALUATIONS OF REGIONAL SULFUR DEPOSITION MODELS.

Investigator(s) and Reference	Model	Predictors Evaluated	Evaluation Tools*	Number of Evaluation Sites and the Region	Evaluation Period
1. Eliassen (1978)	OECD	(a) daily/annual average SO_2 and sulfate air concentrations	(a) r and scatter plots	(a) 50 western and central European sites	1974
		(b) annual average sulfate concentrations in precipitation	(b) analyzed maps	(b) same as (a)	
2. McNaughton (1980)	PNL	(a) monthly average SO_2 and sulfate air concentrations	(a) μ, σ^2, o_{max}/p_{max} comparison, cross sections and histogram of normalized b	(a) 23-38 SURE sites in eastern North America	August and Octobe 1977
		(b) monthly sulfur wet deposition amounts	(b) o/p ratios	(b) 4 MAP3S sites in northeastern U.S.	
3. Voldner et al. (1981)	AES	(a) daily/monthly average sulfate air concentrations	(a) monthly μ, σ, mean b; daily/monthly r and p/o ratios	(a) 19 eastern North American sites	October 1977
		(b) monthly weighted sulfate concentrations in precipitation	(b) same as (a) but only for monthly periods	(b) 15 southeastern Canadian and northeastern U.S. sites	

* Definition of symbols:

r	correlation coefficient
μ	means
b	bias
σ, σ_b	standard deviation and standard deviation of bias
o, o_{max}	observation and maximum observation
p, p_{max}	prediction and maximum prediction
RMSE	root mean squared error
M	model performance error or $(p-o)^2/o^2$

TABLE A.1 CONTINUED

Investigator(s) and Reference	Model	Predictors Evaluated	Evaluation Tools*	Number of Evaluation Sites and the Region	Evaluation Period
4. Bhumralkar (1981)	ENAMAP-1	monthly average SO_2 and sulfate air concentrations	analyzed maps of o and p	54 SURE sites in eastern North America	August and October 1977 and January April, July and October 1978
5. Memorandum of Intent (1982)	AES ASTRAP ENAMAP-1 MEP MOE MONTE CARLO RCDM UMACID	(a) monthly average SO_2 and sulfate air concentrations (b) monthly sulfur wet deposition amounts	(a) linear and logarithmic μ, σ, mean b, σ_b and r of b with p (b) same as (a)	(a) 54 SURE sites in eastern North America (b) 4-13 MAP3S and CANSAP sites	January and July 1978 plus extra polation for 1978
6. Shannon et al. (1982)	AIRSOX ASTRAP RAPT	monthly average SO_2 and sulfate air concentrations	analyzed maps	54 SURE sites in eastern North America	July 1978
7. Voldner and Shannon (1983)	ASTRAP	(b) annual sulfur wet deposition amounts	(b) μ, σ, mean b σ_b,r,r of b with p, p/o ratios % frequency of p within 30, 60 and 90% of o, and 1981 scatter plots	8-14 CANSAP sites in eastern Canada	1980 and 198

TABLE A.1 CONTINUED

Investigator(s) and Reference	Model	Predictors Evaluated	Evaluation Tools*	Number of Evaluation Sites and the Region	Evaluation Period
8. Stewart et al. (1983b)	ENAMAP-1 MIT RTM-II	(a) monthly average SO_2 and sulfate air concentrations	(a) monthly/annual μ, σ, σ/μ, b, σ_b. RMSE, r for (i) all pairs of o and p (ii) pairs with highest 50% of o's. (iii) pair with highest 10% of o's.	(a) 9-54 SURE sites in eastern North America	October 1977 through September 1978
		(b) monthly sulfur wet deposition amounts		(b) 4-7 MAP3S sites in eastern U.S.	
9. Stewart et al. (1983a)	RTM-II	daily average SO_2 and sulfate air concentrations	μ, σ, σ/μ, b, σ_b, RMSE, r	54 SURE sites in eastern North America	July 16-23, 1978
10. Norwegian Meteorological Institute (1984)	MSC-W	(a) annual and 5-year average SO_2 and sulfate air concentrations	(a) r and scatter plots	(a) 27-49 western and central European sites	1978-1982
		(b) annual and 5-year average precipitation weighted sulfate concentration	(b) same as (a)	31-48 western and central European sites	
11. Ruff et al. (1985)	ENAMAP-1 RTM-II UMACID	6-h average SO_2 and sulfate air concentrations	b, RMSE, r	9 SURE sites in eastern North America	January, April, July and October 1978

TABLE A.1 CONTINUED

Investigator (s) and Reference	Model	Predictors Evaluated	Evaluation Tools*	Number of Evaluation Sites and the Region	Evaluation Period
12. Klug and Lupkes (1985)	10 models	(a) annual average SO_2 and sulfate air concentrations	(a) μ, σ, r, m analyzed maps	(a) 26 and 35 western and central European sites for SO_2 and sulfate, respectively	1978
		(b) annual sulfur wet deposition amounts	(b) same as (a)	(b) 24 western and central European sites	
13. Lehmhaus et al. (1986)	MSC-W	(a) monthly/annual average SO_2 and sulfate air concentrations	(a) $\mu, \sigma, r,$ annual scatter plots	(a) 50 central and western European sites	1980
		(b) monthly/annual average precipitation weighted sulfur wet deposition amounts	(b) same as (a)	(b) 36 central and western European sites European sites	

APPENDIX B

STANDARDIZED MODEL INPUT DATA (FIGURES AND TABLES)

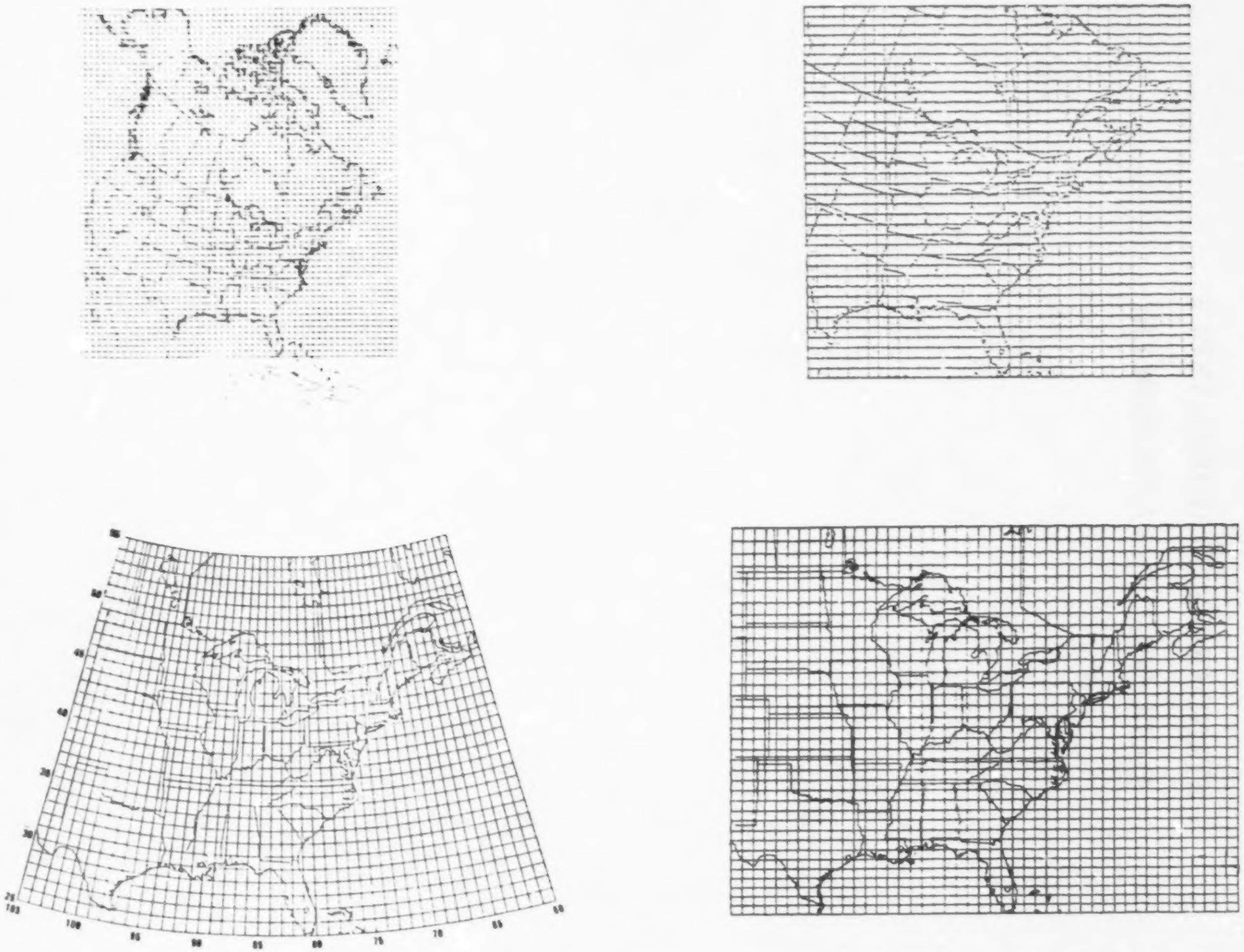


Figure B.1 The four model grid configurations for which model input data have been processed: a) the 52-by-58, 127-km polar stereographic CMC grid, b) the 38-by-37 window of the CMC grid, c) the 45-by-30, 1-degree RELMAP grid, and d) the 45-by-36, 80-km UMACID grid.

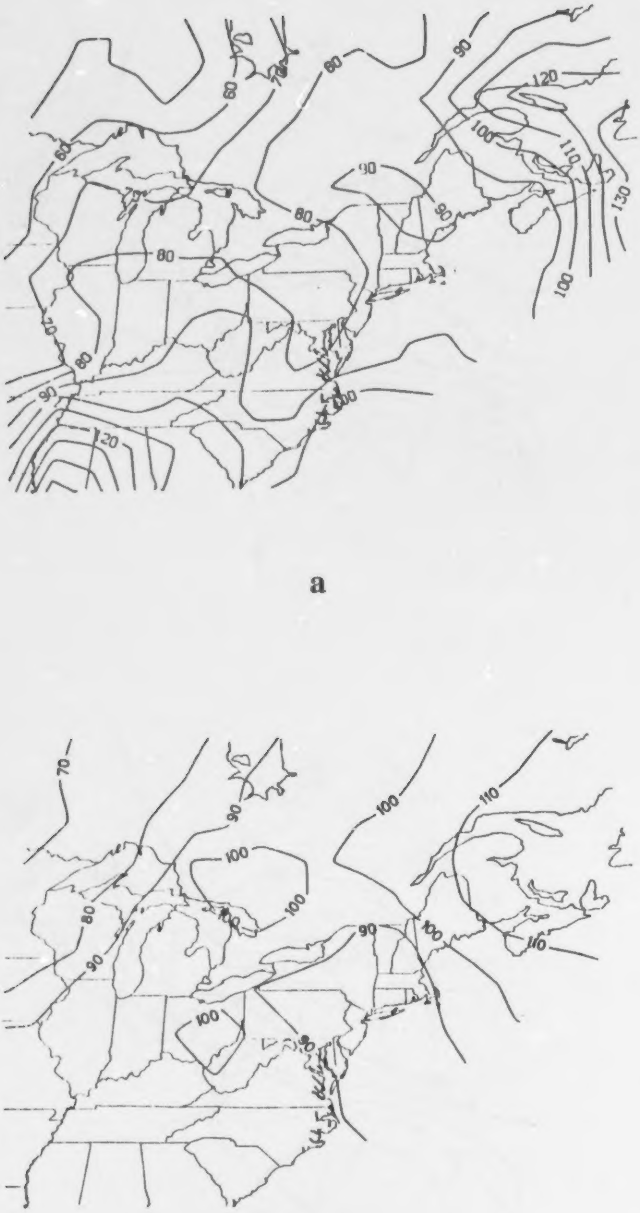


Figure B.2 **Annual** 1980 patterns of precipitation (cm) using a) CMC gridded data and b) data from precipitation chemistry monitoring sites.

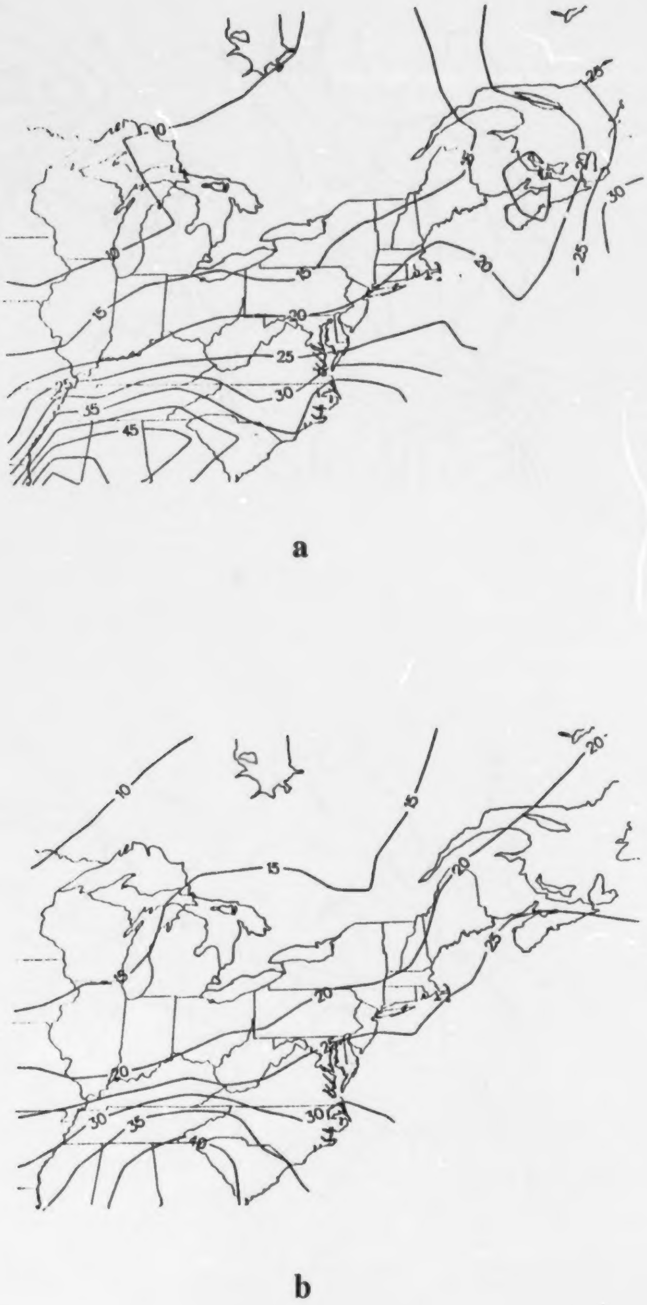


Figure B.3 **Winter** 1980 patterns of precipitation (cm) using a) CMC gridded data and b) data from precipitation chemistry monitoring sites.

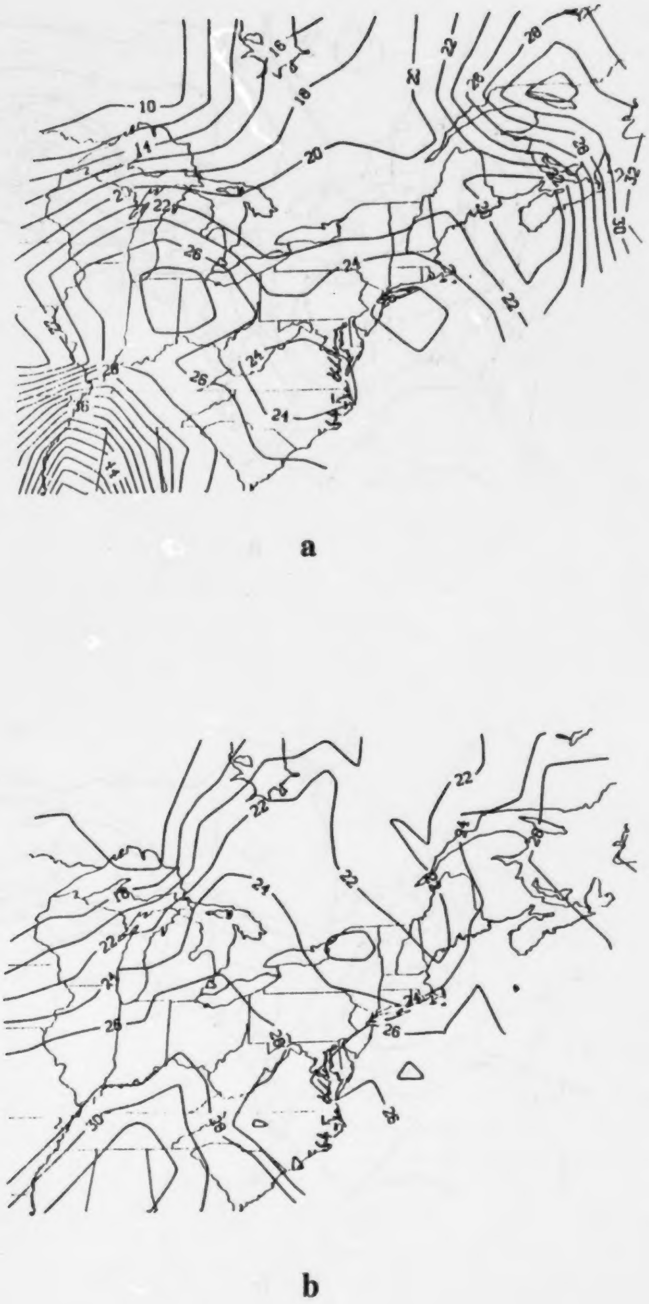


Figure B.4 **Spring** 1980 patterns of precipitation (cm) using a) CMC gridded data and b) data from precipitation chemistry monitoring sites.

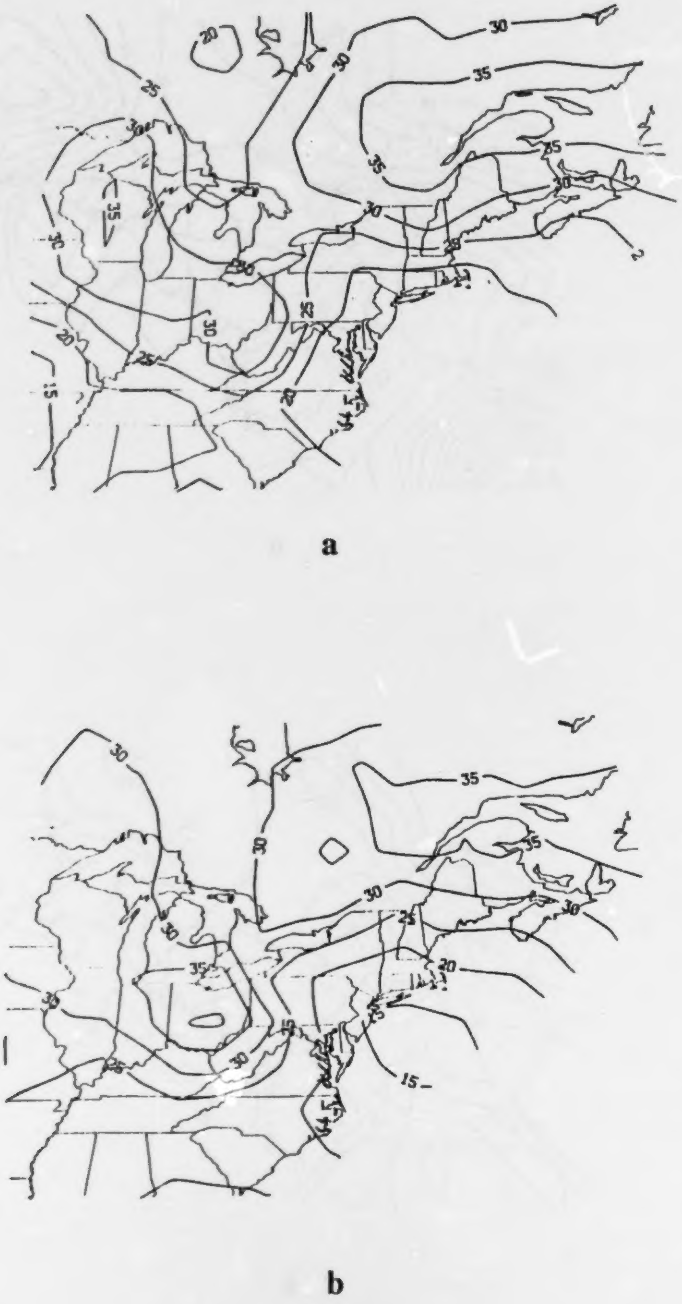
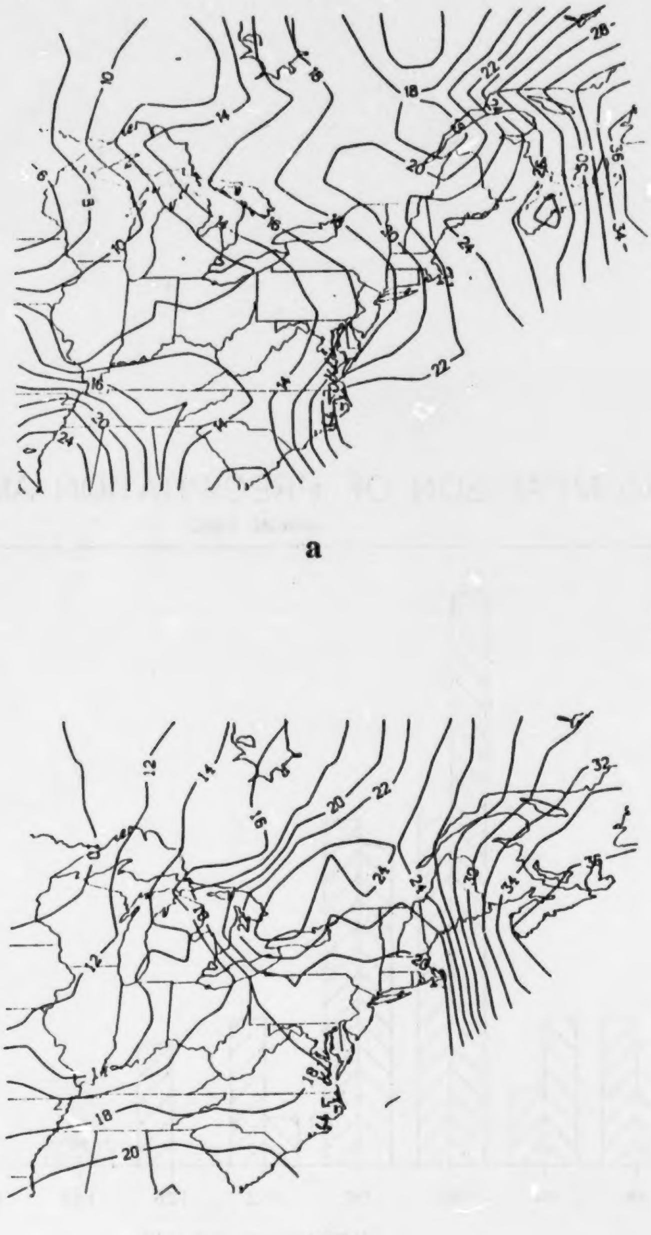


Figure B.5 **Summer** 1980 patterns of precipitation (cm) using a) CMC gridded data and b) data from precipitation chemistry monitoring sites.



a

b

Figure B.6 **Autumn** 1980 patterns of precipitation (cm) using a) CMC gridded data and b) data from precipitation chemistry monitoring sites.

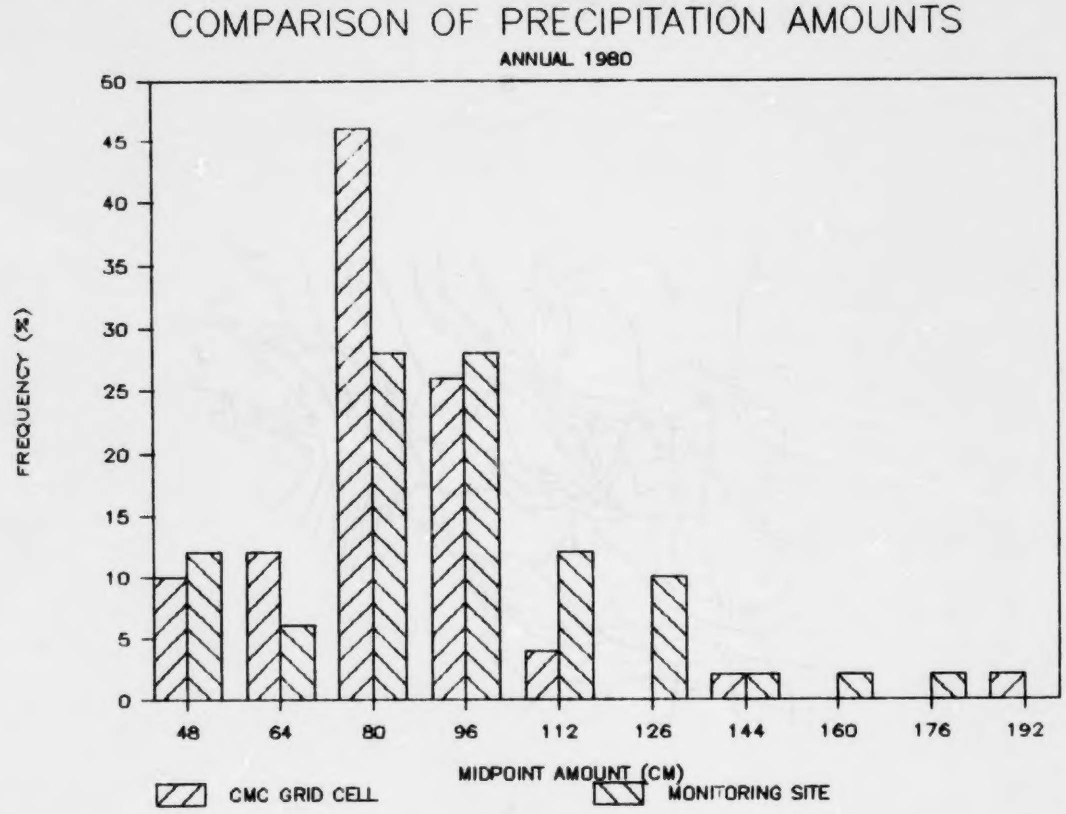


Figure B.7 Histograms of 1980 annual precipitation amounts a) for 127-km grid cells containing precipitation chemistry monitoring sites and b) at the 72 precipitation chemistry monitoring sites across eastern North America.

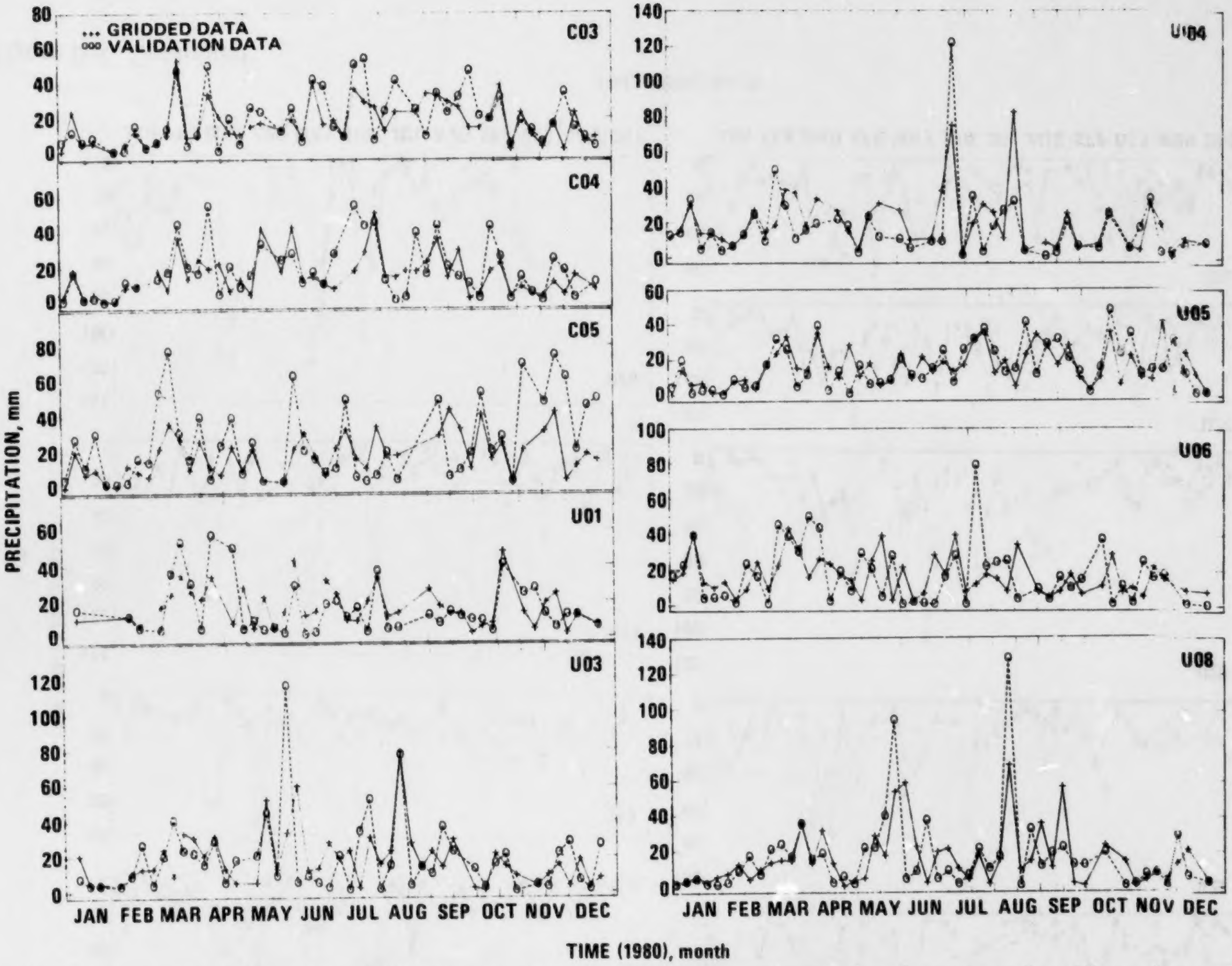


Figure B.8 A comparison of the weekly time series of precipitation amounts (mm) measured at 18 precipitation chemistry monitoring sites (validation data) and calculated for respective 127-km grid cells by the CMC model.

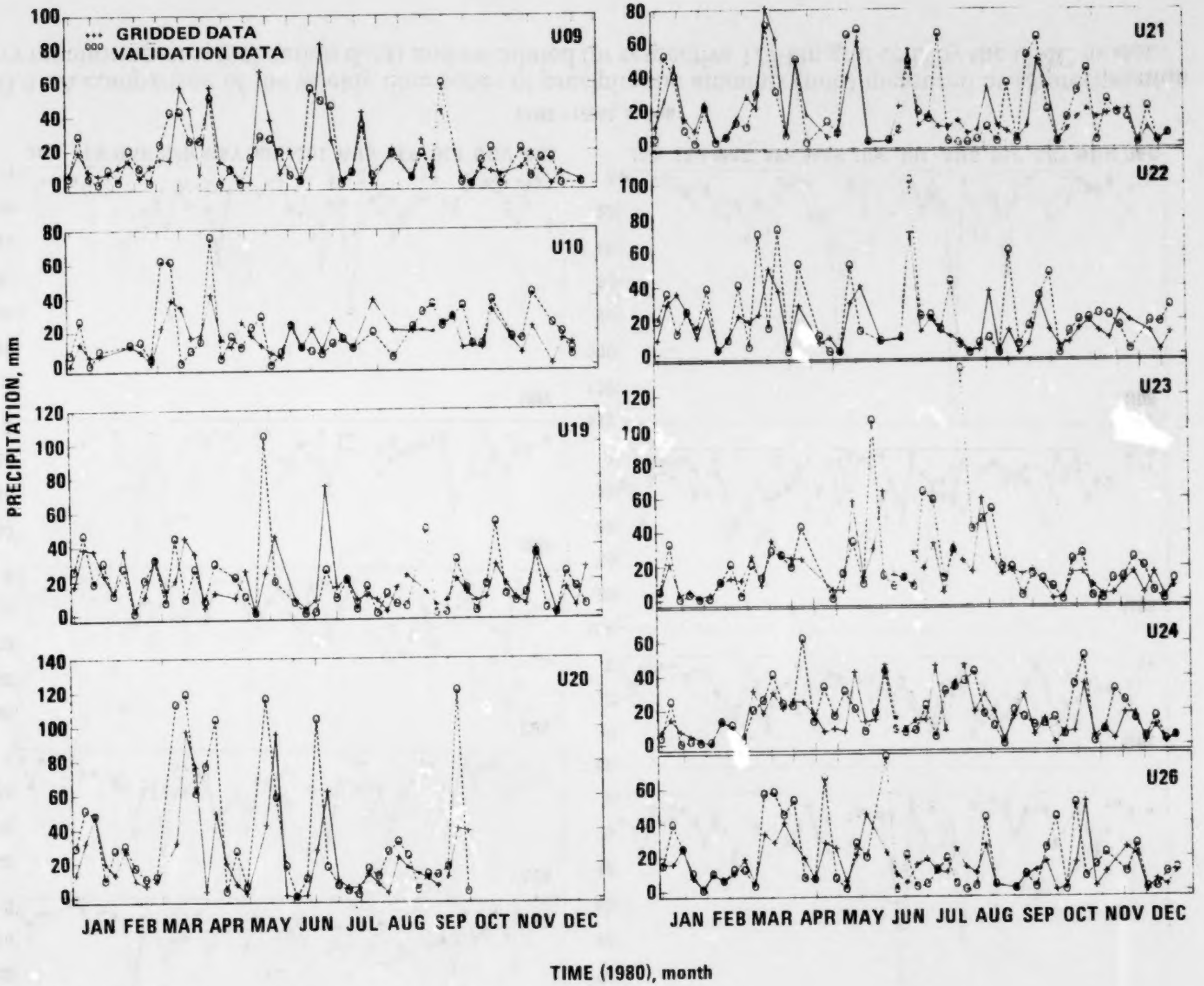


Figure B.8 Continued

TABLE B.1 THE 1980 ANNUAL SULFUR DIOXIDE EMISSION RATES (KILOTONNES PER YEAR) AND PRECISION ESTIMATES FOR CANADA #.

Province	Rate	Precision (%)
Alberta	543.2	±29.3
Nova Scotia	218.0	±14.5
British Columbia	193.3	±17.3
Ontario	1826.5	± 9.3
Manitoba	490.9	±10.3
Prince Edward Is.	5.7	±44.1
New Brunswick	215.0	±14.4
Quebec	1157.0	±14.5
Newfoundland	60.7	±25.7
Saskatchewan	56.8	±42.3
N.W. Territories	2.2	*
Yukon Territory	0.5	*
	4769.8	± 6.1

# (MOI, 1982b)

* No estimate available.

TABLE B.2 THE 1980 ANNUAL SULFUR DIOXIDE EMISSION RATES (KILOTONNES PER YEAR) FOR THE UNITED STATES #.

State	Rate	State	Rate
Alabama	762.8	Nebraska	82.2
Arizona	762.5	Nevada	132.3
Arkansas	78.7	New Hampshire	100.7
California	493.2	New Jersey	291.7
Colorado	109.7	New Mexico	214.2
Connecticut	68.0	New York	787.0
Delaware	103.1	North Carolina	605.0
D. C.	13.9	North Dakota	137.2
Florida	1111.1	Ohio	2612.4
Georgia	807.6	Oklahoma	92.3
Idaho	50.4	Oregon	50.5
Illinois	1392.9	Pennsylvania	1724.7
Indiana	1744.0	Rhode Island	13.0
Iowa	434.5	South Carolina	325.1
Kansas	209.7	South Dakota	68.7
Kentucky	1065.9	Tennessee	1049.6
Louisianna	468.2	Texas	1204.3
Maine	145.9	Utah	119.3
Maryland	296.7	Vermont	8.5
Massachusetts	254.4	Virginia	357.0
Michigan	813.7	Washington	295.2
Minnesota	248.6	West Virginia	1041.1
Mississippi	274.6	Wisconsin	641.2
Missouri	1172.5	Wyoming	228.1
Montana	151.7		
		Total	25154.4

# Version 2.0 of the National Acid Precipitation Assessment Program (NAPAP) emission inventory (unpublished).

TABLE B.3 UNITED STATES AND CANADIAN STACK HEIGHT FREQUENCY DISTRIBUTION.

Physical Stack Height Interval (m)	Canadian Point Sources		United States Point Sources	
	Number	SO_2 Emissions (ktonnes/yr)	Number	SO_2 Emissions (ktonnes/yr)
0 - 50	450	543.7	100367	3481.2
51 - 100	208	636.3	5728	4972.5
101 - 150	45	752.7	374	2754.0
151 - 200	21	1011.5	281	4520.7
201 - 250	1	16.3	111	2294.1
251 - 300	3	249.3	46	1089.7
301 - 350	0	0.0	55	1818.3
351 - 400	2	746.3	6	237.8
Missing Heights	0	0.0	69591	2383.4
Total	730	3956.1	176559	23551.7

APPENDIX C

SCREENING AND CALCULATION PROCEDURES FOR THE MODEL EVALUATION DATA SET

Excerpts from the Final Report of Unified Deposition Data Base Committee(1985) reprinted with permission of the primary author.

C.1 Concentration and Deposition Calculation Procedures

All seasonal and annual concentrations are precipitation-weighted average values defined as

$$C = \frac{\sum_{i=1}^{N} C_i \, P_i}{\sum_{i=1}^{N} P_i}$$

where

C = the seasonal/annual weighted concentration (μmoles/l);

C_i = the daily/weekly/monthly concentration (μmoles/l) measured during collection period i;

P_i = the precipitation depth (cm) recorded during collection period i; and

N= number of simultaneous valid measurements of both precipitation and concentration during the season/year.

Valid sample criteria applicable to the annual and seasonal summaries are:

1) All sampling periods for which it is known that no precipitation occurred are considered valid sample periods. This applies mainly to weekly and monthly sampling protocols. For event and daily sampling protocols the absence of a sample record for a day implies that no precipitation occurred (this assumption is not always valid for the earlier data).

2) The wet deposition sample must be a wet-only sample. All samples identified as bulk, partially bulk, or undefined are invalid.

3) Wet deposition samples that have insufficient precipitation to complete a chemical analysis for a specific ion species are invalid for that specific ion species. Event and daily samples are most likely to have this occur.

4) The concentration of an ion species accompanied by a comment code designating the measurement to be "suspect" or "invalid" is declared an invalid sample. Deletion of the ion species concentration by network personnel for the same reason has the same effect.

5) The actual sampling period for a wet deposition sample must be close to the network's protocol sampling period.

All seasonal/annual deposition values are calculated by multiplying the precipitation-weighted average concentrations by the total precipitation amounts for the season/year (P_t) defined as

$$D = C P_t$$

where

$$P_t = \sum_{i=1}^{N} P_i$$

and N is the number of collection periods with precipitation amounts for the season/year. Valid sample results designated as below the detection limit are included in the average value as one-half of the detection limit.

The precipitation depth is generally obtained from a colocated standard gauge measurement. At some locations, however, there may be a systematic overcatch problem (i.e. the collector depth is greater than the standard gauge depth. This type of situation is relatively rare but, when it does occur, a malfunctioning standard gauge is normally the source of the problem. For these cases, the gauge and collector depth data are carefully examined and the deposition is calculated using both the gauge depth as well as the maximum of gauge or collector depths (whichever is greater during each collection period). If the deposition calculated using the maximum of gauge or collector depths is more than 10% greater than the deposition calculated using the gauge depth alone and if there is no obvious problem with either the gauge or collector, then the concentration or deposition is calculated using the greater of the two values, i.e.,

$$C = \frac{\sum_{i=1}^{N} C_i \text{ (}P_i \text{ or } P_{ic}\text{, whichever is greater)}}{\sum_{i=1}^{N} \text{(}P_i \text{ or } P_{ic}\text{, whichever is greater)}}$$

and

$$D = C \sum_{i=1}^{N} \text{(}P_i \text{ or } P_{ic}\text{, whichever is greater)}$$

where

P_i = the standard gauge depth (cm) and

P_{ic} = the collector depth (cm).

Precipitation chemistry data are often missing from one or more collection periods. In these cases the annual/seasonal deposition is handled by multiplying the precipitation-weighted concentration (using the available data) during the year/season by the measured total precipitation depth (P_t) during the year/season.

The precipitation depth data are collected at the precipitation chemistry monitoring site. If depth data are missing from the standard gauge for a particular sampling period, then different procedures have historically been used in Canada and the United States. In Canada, interpolated gauge amounts from surrounding weather service stations are generally used, while in the United States collector depths are used.

C.2 Definition of Annual and Climatological Seasons

Preparation of annual and seasonal data summaries requires specific definitions for annual and climatological seasons. Annual periods are here selected to coincide as closely as possible to the calendar year. Seasonal periods are defined to coincide as closely as possible to the climatological seasons, defined here as: winter - January, February and March; spring - April, May and June; summer - July, August and September; autumn - October, November and December.

The fact that some cumulative networks use collection periods that do not necessarily coincide with the preferred annual and climatological seasonal periods based on calendar months presents a problem. For example, NADP follows a weekly collection period starting on Tuesday. Entire collection periods must be allocated to an annual or seasonal period either on a fractional basis or a fractional precipitation occurrence basis. The latter requires knowledge of when precipitation occurred within a single collection period. Since this information is not available for all networks, the allocation of collection periods is computed here on a fractional time basis.

The procedure used here for the fractional time allocation of samples to an annual or seasonal period requires that at least fifty percent of the time covered by a collection period be contained in the annual or seasonal period. This applies to both the beginning and end of an annual or seasonal period. For each summary the first and the last date of the period summarized is given in the data tables.

C.3 Screening Criteria

Data quality is an important consideration in any wet deposition summary process. Aspects of data quality include site representativeness, network operation and sample analysis procedures, screening of individual samples on the basis of available sample comment codes, and data completeness measures for the period being summarized. Network procedures and the screening process applied to individual samples are discussed in Appendix A of the UDDBC Report. Data quality levels discussed below are based on only those samples that meet the screening criteria described in this Appendix.

C.3.1 Site Representation

Each valid mean concentration and deposition amount is assigned an overall data quality level ranging from "1" to "3", where level "1" is the highest level. This overall data quality level is based on a *site representativeness rating* and a *data completeness rating*. Regionally unrepresentative sites, those sites meeting one or more of the criteria in Table C.1, are considered to be unrepresentative for monitoring regional scale acid deposition. These sites receive a rating of "3" unless an examination of the historical data from the site, or the judgement of those familiar with the site, indicated that the local interferences were small or insignificant.

TABLE C.1 CRITERIA FOR UNREPRESENTATIVE MONITORING SITES.

1. There is a continuous industrial source, a town or suburban area located within 10 km or
2. There is a major point source (greater than 10,000 tons sulfur dioxide or nitric oxide per year) located within 50 km or the sum of point source emissions within 50 km is greater than 10,000 tons per year or
3. There is a surface pollutant storage facility (e.g. salt pile) located within 100 m or
4. There are transportation sources, furnaces or incinerators located within 100 m or
5. There is cultivation or other agriculture activity within 500 m or
6. Buildings, trees, etc., impinge on the cone defined by a 45 degree angle above the horizontal plane and centered at the site (30 degrees is considered optimal but 45 degrees is the highest acceptable angle or
7. The local area is known to be dusty due to poor ground cover.

On the other hand, sites with a rating of "1" are not exposed to any of the interferences listed in Table C.1 and are considered to be regionally representative. However, there is a large number of sites which fall into neither the "3" nor "1" categories, and for which evaluation of site representativeness is difficult. The comments of individual network operators, analyses of the historical data, or site audits were used to distinguish further between type "2(a)" and "2(b)" sites. In the former case, those familiar with the site or with historical data from the site, have judged that the local interferences are probably small or insignificant and the sites are potentially representative. On the other hand, "2(b)" sites have more significant problems associated with them, but those familiar with their surroundings or historical data did not feel that they should clearly go into the "3" category. They were nevertheless judged to be potentially unrepresentative and data from them should be regarded with somewhat less confidence than data from "2(a)" sites.

It should be noted that detailed site evaluations (based on both the nature of the surroundings and historical data characteristics) are presently underway in a number of networks. Therefore, the ratings used in the ISDME, although based on the best information available at the time, should be considered as tentative and subject to possible revision.

C.3.2 Data Completeness

Each annual and seasonal average value is assigned one of three data completeness ratings. A value not meeting the lowest rating requirements is excluded from the data tables. Five data completeness measures are used to determine the data completeness rating. A sixth measure, percent sea salt correction, is used for sulfate values for sites within 100 km or the coast. All six criteria must be met before an individual sulfate or nitrate value is assigned a given data completeness rating. How these measures are used to determine the data completeness rating is given in Tables C.2 and C.3. The abbreviations of the data completeness measures in the these tables are defined as follows:

% PCL - Percent Precipitation Coverage Length is defined as the percent of the summary period for which information on whether or not precipitation is available:

$$\% \text{PCL} = 100\,(\text{SPL} - \text{NDDPP})/\text{SPL}$$

where

SPL = the number of days in the summary period and

NDPP = the number of days when it is not known if precipitation occurred.

If precipitation is known to have occurred during a particular sampling period but no measurement of the amount is available then no knowledge of precipitation is assumed. Total precipitation depth P_t is the amount of precipitation occurring during the period of precipitation coverage. This data completeness measure does not include any consideration of the availability of a valid precipitation chemistry sample.

% VSL - Percent Valid Sample Length is the percent of the days in the summary period that are associated with valid sample periods:

$$\% \text{VSL} = 100\,(\text{NDNP} + \text{NDVCMP})/\text{SPL}$$

where

NDNP = the number of days in the summary period associated with sample periods during which no precipitation occurs.

NDVCMP = the number of days in the summary period with valid sample component measurement on measured precipitation sample.

% COL EFF - Percent Collection Efficiency is the ratio (converted to a percent) of the total sample volume (converted to a gauge depth) to the total precipitation depth:

$$\% \text{COL EFF} = 100\,(\text{EPDVC}/\text{ERGVC})$$

where

EPDVC = the sum of depths calculated from the sample volume for qualifying samples and

ERGVC = the sum of the standard gauge depths for qualifying samples.

Qualifying samples are those that have both (a) a collocated standard gauge and sample volume measurement available and (b) a valid sample component measurement. The collection efficiency requirement for Canadian winter average values has been relaxed somewhat compared to other seasons due to the generally poorer collector performance for snow sampling. If the 80% requirement for other seasons (level 1) is applied to the winter months when a very large percentage of the precipitation in Canada is in the form of snow, then the data loss is quite large. It is felt that a lower percentage could be accepted for winter because the problems are primarily due to undercatch of snow. The sample collected is expected, therefore, to be representative. The Canadian annual collection efficiency requirement is also relaxed somewhat as a consequence.

% SEASALT - Percent Sea Salt Correction is the percent of the average sulfate concentration that is estimated to be due to sea salt:

$$\% \text{ SEASALT} = 100\,(\text{XBAR} - \text{ABARC})/\text{XBAR}$$

where

XBAR = the arithmetic mean of sulfate concentrations from uncorrected valid sulfate samples and

XBARC = the arithmetic mean of sulfate concentrations from sea salt corrected valid sulfate samples.

The sea salt correction uses sodium or magnesium as tracers of sea salt, as described in Appendix D of the UDDBC Report.

TABLE C.2 DATA COMPLETENESS RATING CRITERIA FOR SEASONAL SULFATE VALUES.

Data Completeness Measure	Level 1	Level 2	Level 3
% PCL	≥95%	≥90%	≥90%
% VSL	≥80%(60%)	≥70%(50%)	≥60%(30%)*
% COL EFF			
Winter	≥75%(40%)*	≥50%(30%)*	≥40%(20%)*
Spring Summer Fall	≥75%	≥50%	≥40%
% SEASALT (sulfate)	≤25%	≤50%	≤75%

* The values within the parentheses apply to Canadian data only.

TABLE C.3 DATA COMPLETENESS RATING CRITERIA FOR ANNUAL SULFATE VALUES.

Data Completeness Measure	Level 1	Level 2	Level 3
% PCL			
Annual	≥95%	≥90%	≥90%
Each Quarter^	≥75%	≥60%	≥50%
% VSL			
Annual	≥80%	≥70%	≥60%
Each Quarter^	≥70%	≥60%	≥50%
% COL EFF			
Annual	≥75%(65%)*	≥50%(45%)*	≥40%(35%)*
Winter	≥75%(40%)*	≥50%(30%)*	≥40%(20%)*
Spring Summer Autumn	≥75%	≥50%	≥40%
% SEASALT (sulfate)	≤25%	≤50%	≤75%

^ An annual summary must not only meet the annual data completeness criteria, but also the quarterly criteria for each quarter. The addition of quarterly criteria for % TP, VSL and VSMP to the annual criteria is to insure that adequate data from each quarter is present in the annual summary. The quarterly criteria for % TP, VSL and VSMP in Table C.3 are relaxed from the seasonal criteria in Table C.2 since the emphasis is on insuring adequate data for an annual summary, and not reporting a quarterly summary.

* The values within the parentheses apply to Canadian data only.

C.4 Assessment of Uncertainties in Deposition Amounts

Errors in deposition calculated from precipitation chemistry measurements can arise from several sources such as actual sampler design, errors due to laboratory procedures, field related errors caused by sampling periods, poor catch, contamination and missing samples. These errors are difficult to quantify and may vary from network to network. The magnitude of these errors has been investigated here by

1) Reviewing studies aimed at quantifying network bias as well as accuracy and precision of measurements;

2) Statistically analyzing the ramifications of missing sampling periods; and

3) Comparing observed deposition amounts for 1980-1983 at colocated or closely located samplers.

Vet et al. (1981) intercompared the performance of CANSAP and APIOS collectors at five sites in Ontario. Samples were collected on a monthly basis from spring 1979 to spring 1980. They found that the CANSAP data averaged 20 percent higher than APIOS sulphate concentrations. De Pina et al. (1980) intercompared samplers used in the CANSAP, MAP3S, NADP and EPRI networks during the period July 4, 1978 to April 30, 1979. Since NADP and EPRI used the same sampler type it is referred to here as EPRI.

Detailed analysis of these data by Vet and Onloch 1981 (see also Vet and Reid, 1982) revealed that the mean values for deposition ratios of sulfate calculated by event was 1.24 ± 0.47 for CANSAP to MAP3S and 1.16 ± 0.38 for CANSAP to EPRI. Over the entire period CANSAP overestimated sulfur deposition compared to the other samplers by 20 and 15 percent, respectively. The differences may be somewhat lower in 1980, since the CANSAP operational protocol changed. Bias and precision in the EPRI daily network for 1979-1980 were estimated by Topal et al. (1984). They detected a consistent anion-cation inbalance, larger than observed in the NADP network (Stensland and Bowersox, 1984), but were unable to trace its origin. The relative annual precision in concentration and precipitation based on pooled data from colocated samplers at nine sites were 6% for sulfate concentration and 5% for precipitation. To discount outliers, the annual precision was defined as the median value of the colocated sample differences (ACD) and the relative precision was ACD per median value of concentration or precipitation.

A recent study of the APIOS networks (Lusis, personal communication) found the reproducibility of sulfur concentration in precipitation with wet-only Sangamo sampler (used in the cumulative network) over a 28 day period was at least 88%. They experienced less than 10% evaporative losses of precipitation and estimated accuracy of deposition to better than 10%. The annual wet deposition measured by the 28d cumulative Sangamo sampler was about 10% higher than thatdetermined from the daily Aerochem Metric samplers. Thus, network differences can easily account for more than 10% differences in estimated deposition amounts.

Uncertainties due to missing data in 1980 were estimated for the NADP, APN and CANSAP networks. Weekly concentration amounts were removed randomly throughout winter and summer seasons. In the Canadian data, for which precipitation is always available, the seasonal deposition was recalculated from the remaining concentration data by scaling with seasonal precipitation. For the NADP data, where precipitation coverage may deviate from 100%, the new seasonal deposition was computed in two ways, i.e. as for the APN and from periods, when both concentration and precipitation were available. The latter approach creates a consistent underestimation. When 1, 2, 3 and 4 weeks of data were missing, the coefficient of variation was about ± 4, ± 6, ± 7, and ± 9% in for the APN data, while a coefficient of variation and an underestimation of 8, 16, 24 and 30% were observed in the United States data. Absolute maximum errors at the individual sites for any reason were 10, 20, 30, 40%, respectively, for the APN network.

Missing periods - The seasonal/annual precipitation weighted mean concentrations are determined from simultaneous valid measurements (n) of precipitation and concentrations. This concentration is multiplied by the total precipitation depth measured for the season/year to yield the seasonal/annual deposition. For the Canadian networks and frequently for the NADP network, estimates of precipitation are available for each collection period (N) within the season/year, while in the MAP3S and EPRI networks precipitation is not always recorded (i.e. number of collection periods <N). Thus, compensation for missing measurements in the Canadian data and generally in the NADP data introduces random error (i.e. over or under prediction) in the estimated deposition. If concentration and precipitation measurements are simultaneously missing in the MAP3S or EPRI data bases, the resulting deposition is underestimated. However, N_1 is generally less than N but greater than n. Since missing periods were not recorded, in the EPRI-SURE network, deposition estimates may be underestimated.

Difference in deposition estimates from colocated or closely located samplers were computed for 1980 and 1982. The "colocated" samplers included samplers from the same network as well as from different networks. The data were obtained from the ADS summary.

The annual and seasonal results for 1980 are provided in Table C.4. The table gives the mean and maximum absolute difference between pairs of sites with observations rated either both a "1" or a "1" and "2", within 1 km and 90 km of each other. The latter distance was chosen, since it corresponds to the spatial resolution of most of the ISDME models. Thus, the observed differences would include the local variability in the data, which cannot be captured by the models. Similarly, Table C.5 gives the annual results for 1981 and 1982 (seasonal deposition amounts were not available from ADS). Due to higher network density, the upper distance was reduced to about 25 km.

As seen from the tables, differences between pairs of observations both classified as "1" and taken within 1 km of each other amount to 10%. The average difference increases to over 20% for pairs of observations, one of which is rated as a "2", within 25 km of each other. It is interesting to note that differences greater than 50% are observed in the 1980 seasonal data. Thus, ± 20% can been seen as a conservative estimate of errors and local variability in the observed deposition amounts.

APPENDIX D

COMPARISON OF PREDICTIONS AND OBSERVATIONS OF SULFUR WET DEPOSITION AT THE ISDME SITES

TABLE D.1 COMPARISON OF THE OBSERVATIONS AND PREDICTIONS OF ANNUAL 1980 SULFUR WET DEPOSITION (KG S/HA) AT THE ISDME EVALUATION SITES.

ISDME CODE	VALIDATION SITE	MODELS										
		TVA	AES	SIMPMOD	STATMOD	MOE	SERTAD	RTM-II	RIVAD	ASTRAP	UMACID	RELMAP
C01	2.05	1.17	0.80	0.96	0.54	2.03	3.29	1.70	1.52	1.40	0.62	0.27
C03	9.00	6.38	7.97	6.76	3.99	7.01	10.18	4.39	4.19	10.46	5.28	9.01
C04	11.67	12.67	9.28	11.20	7.61	11.51	11.79	4.93	5.57	16.07	9.46	14.00
C05	5.83	3.43	4.83	4.19	2.51	2.42	6.13	4.06	2.73	6.57	2.79	6.10
C11	.	0.87	0.42	.	.	1.18	3.55	.	0.39	0.84	.	.
C12	.	1.06	0.62	.	.	1.76	3.91	1.68	1.25	1.22	.	.
C13	.	1.78	0.80	1.43	0.56	2.93	5.14	2.02	1.54	2.84	.	0.88
C14	11.03	8.36	7.67	7.29	4.16	8.46	10.89	4.13	4.24	11.10	7.89	8.58
C15	15.77	11.91	9.96	12.49	11.17	10.83	10.34	5.14	5.24	14.99	8.81	11.07
C16	12.20	8.81	8.84	8.09	4.21	8.09	12.24	4.07	4.53	13.82	8.76	8.95
C17	13.50	7.46	9.81	7.46	3.76	6.79	11.25	4.26	4.74	13.18	8.93	8.67
C18	.	0.53	0.16	0.51	0.18	0.40	3.19	.	0.00	0.53	.	.
C19	2.77	1.18	0.80	1.17	0.55	1.70	5.16	.	0.91	2.03	.	0.72
C20	5.13	2.67	2.96	2.74	1.69	3.55	7.65	3.20	2.50	5.41	0.77	3.81
C22	3.53	0.75	0.54	0.93	0.63	0.70	6.13	.	0.73	1.36	.	0.53
C23	9.20	2.67	2.93	2.94	2.03	3.15	7.36	3.52	2.65	5.87	1.86	3.15
C24	9.13	3.16	4.08	3.76	2.31	3.51	7.64	3.85	3.03	7.17	2.41	4.64
C26	.	1.17	1.64	1.58	1.54	0.97	7.84	.	1.16	2.78	.	.
C33	.	5.63	.	5.95	3.30	5.89	.	.	.	10.97	.	6.18
C34	.	2.64	3.27	2.86	1.65	3.37	6.93	3.20	2.61	6.27	.	2.91
U01	5.57	10.72	11.23	12.23	6.44	7.90	10.97	4.91	5.47	16.08	6.13	14.98
U02	.	9.31	9.52	9.26	8.66	4.26	9.35	4.55	3.42	7.21	6.32	8.61
U03	8.63	10.20	9.43	9.21	7.34	8.60	10.84	5.19	5.26	11.38	6.05	10.95
U04	5.33	12.83	15.33	12.71	8.37	6.98	11.12	6.06	5.12	10.63	11.69	15.28
U05	7.47	6.49	10.47	6.78	3.65	5.75	9.97	4.91	4.66	12.11	5.21	7.81
U06	7.03	11.38	13.45	12.61	6.75	6.17	9.57	5.49	4.88	10.24	6.48	20.08
U07	.	3.92	4.04	4.02	3.18	1.02	6.85	4.36	1.88	3.71	.	3.13
U08	9.93	10.19	7.72	8.26	6.45	7.97	8.97	4.78	4.53	9.68	5.22	8.56
U09	.	12.12	8.66	11.47	10.33	6.25	7.49	4.58	3.90	8.25	4.65	6.47
U10	7.13	5.95	9.52	6.57	3.86	4.97	9.08	5.16	4.52	11.39	3.70	10.04
U11	.	4.22	6.60	4.97	3.20	4.23	9.20	4.60	3.87	8.93	3.52	6.17
U12	.	4.59	4.13	3.64	2.07	6.64	7.38	3.21	3.34	6.29	2.12	2.82
U13	.	9.00	8.49	8.04	5.79	8.61	10.05	4.71	5.02	10.01	4.92	10.57
U14	.	6.12	5.29	4.82	3.11	7.26	8.88	3.72	3.93	7.37	3.33	4.08
U15	.	1.72	1.44	1.54	0.85	2.75	3.45	2.02	2.63	2.01	0.74	0.59
U16	.	3.98	5.68	4.11	4.84	1.93	9.80	4.47	2.42	3.84	.	3.21
U17	.	11.27	10.75	11.26	5.57	9.04	10.92	4.80	5.76	16.47	8.08	15.13
U18	.	11.82	11.69	14.08	9.36	7.02	8.74	4.97	4.90	15.80	.	16.70
U19	6.23	7.39	10.38	8.25	5.39	2.67	7.61	4.68	3.23	7.16	4.68	11.41
U20	.	8.51	10.94	8.83	6.52	3.40	8.73	4.66	3.90	7.19	7.58	12.29
U21	10.07	8.97	10.61	9.28	6.96	3.44	8.45	4.68	3.66	8.65	5.56	10.85
U22	6.23	7.41	8.49	7.91	4.76	2.57	7.24	4.32	3.23	6.59	3.43	8.78
U23	14.40	12.84	10.98	10.94	8.71	10.23	10.90	5.97	6.02	14.38	8.72	13.33

TABLE D.1 CONTINUED

ISDME CODE	VALIDATION SITE	MODELS										
		TVA	AES	SIMPMOD	STATMOD	MOE	SERTAD	RTM-II	RIVAD	ASTRAP	UMACID	RELMAP
U24	9.25	15.65	13.85	16.30	10.13	11.00	12.45	5.97	6.29	23.14	10.70	17.46
U25	.	3.28	2.75	2.65	1.92	5.02	5.88	2.71	2.64	3.96	2.26	1.28
U26	6.93	9.83	12.12	12.68	8.11	4.45	8.27	4.83	4.07	10.77	6.24	15.27
U40	.	3.15	3.93	3.72	2.14	3.58	7.66	3.75	3.12	6.90	2.21	4.84
U41	.	8.77	9.48	10.11	6.36	6.09	9.35	4.60	4.63	13.91	4.21	11.67
U42	8.33	10.06	10.27	10.53	5.42	7.98	7.84	4.62	5.40	15.59	7.18	13.95
U43	6.23	9.25	10.88	12.41	9.19	5.15	8.18	4.57	3.93	13.17	4.27	16.37
U44	14.53	17.96	18.94	19.48	13.79	10.30	11.60	6.58	6.60	17.65	12.89	36.71
U45	11.00	14.88	11.67	12.84	9.84	12.40	11.37	6.25	6.31	16.28	8.82	16.34
U46	13.47	19.63	15.00	15.58	13.33	12.56	12.19	7.30	6.91	22.59	11.28	29.80
U47	.	13.64	10.75	11.73	9.32	8.62	10.35	5.89	6.12	16.67	8.88	15.28
U48	8.67	13.39	9.74	11.78	9.93	7.15	9.15	5.07	4.74	11.42	6.26	11.47
U49	.	8.88	8.34	7.57	7.09	8.44	9.20	4.64	4.54	10.23	5.60	9.70
U50	.	5.80	6.19	5.39	4.75	5.92	5.81	3.80	3.81	5.61	.	5.55
U51	.	5.02	4.62	4.17	3.24	6.42	7.33	3.64	3.79	5.51	.	4.14
U52	.	1.60	1.23	1.26	0.77	2.66	3.54	1.84	1.82	1.95	0.61	0.49
U53	.	2.66	4.05	2.71	2.47	2.23	4.66	2.92	2.11	2.11	.	1.87
U54	.	6.32	8.00	6.10	6.15	3.84	8.29	3.47	2.94	4.30	.	4.53
U55	.	9.86	11.20	10.46	8.39	4.61	8.93	4.85	4.60	9.04	5.71	13.58
U56	.	11.67	12.59	11.34	8.78	6.14	9.41	5.53	5.24	10.23	.	18.06
U57	.	8.16	9.37	8.70	8.18	3.98	9.18	4.66	3.22	8.61	.	13.88
U58	.	9.47	10.55	9.56	7.43	4.21	9.04	4.86	4.21	9.48	6.81	13.45
U59	.	6.45	7.97	6.78	5.18	1.80	7.58	4.08	2.61	6.35	.	7.26

TABLE D.2 COMPARISON OF THE OBSERVATIONS AND PREDICTIONS OF WINTER 1980 SULFUR WET DEPOSITION (KG S/HA) AT THE ISDME EVALUATION SITES.

		MODELS										
ISDME CODE	VALIDATION SITE	TVA	AES	SIMPMOD	STATMOD	MOE	SERTAD	RTM-II	RIVAD	ASTRAP	UMACID	RELMAP
C01	0.32	0.22	0.16	0.30	0.11	0.50	0.23	0.24	0.13	0.28	0.14	0.03
C03	0.83	1.04	1.37	1.94	0.78	1.15	1.28	0.59	0.55	1.83	1.01	0.91
C04	2.38	2.26	1.74	3.78	1.89	1.63	2.71	0.90	0.84	3.13	1.84	1.79
C05	1.23	0.53	0.78	1.26	0.55	0.18	0.44	0.56	0.28	1.11	0.46	0.81
C11	.	0.16	0.07	.	.	0.34	0.22	.	0.03	0.16	.	.
C12	.	0.20	0.12	.	.	0.48	0.27	0.22	0.19	0.26	.	.
C13	0.57	0.32	0.29	0.40	0.10	0.83	0.40	0.33	0.36	0.64	.	0.45
C14	1.93	1.41	1.41	2.37	0.91	1.29	1.50	0.60	0.56	2.16	1.45	1.25
C15	3.20	2.07	2.38	3.89	2.29	1.80	1.74	0.96	0.87	2.75	1.83	1.45
C16	2.23	1.46	1.36	2.55	0.83	1.19	2.00	0.61	0.68	2.54	1.95	1.58
C17	2.57	1.19	1.63	2.33	0.80	1.03	1.70	0.67	0.61	2.36	1.49	1.30
C18	.	0.09	0.01	0.16	0.05	0.08	0.17	.	0.00	0.11	.	.
C19	0.27	0.19	0.12	0.35	0.11	0.36	0.30	.	0.12	0.42	.	0.10
C20	0.60	0.40	0.54	0.74	0.29	0.73	0.69	0.51	0.36	0.98	0.05	0.30
C22	0.67	0.12	0.10	0.28	0.17	0.06	0.22	.	0.07	0.24	.	0.12
C23	1.37	0.41	0.35	0.88	0.38	0.30	0.46	0.48	0.26	0.92	0.28	0.39
C24	1.00	0.48	0.50	1.13	0.47	0.26	0.53	0.54	0.29	1.18	0.32	0.56
C26	.	0.17	0.40	0.45	0.35	0.07	0.32	.	0.13	0.41	.	.
C33	.	0.90	.	1.81	0.69	0.91	.	.	.	1.93	.	0.74
C34	.	0.40	0.56	0.83	0.25	0.52	0.60	0.44	0.31	1.06	.	0.28
U01	0.90	1.67	2.52	3.76	1.52	1.12	2.18	0.91	0.90	2.92	1.37	2.45
U02	2.50	1.76	3.71	2.61	3.33	1.25	1.41	1.50	0.94	2.28	1.90	1.27
U03	1.20	1.72	2.34	2.97	1.94	1.55	1.56	1.09	0.93	2.30	1.30	1.31
U04	0.80	2.33	5.33	4.54	2.62	1.37	2.39	1.63	1.27	2.25	3.23	2.74
U05	0.67	1.06	1.68	2.16	0.86	0.80	1.42	0.75	0.58	2.14	0.71	1.05
U06	1.33	2.04	5.05	4.49	2.55	1.08	2.12	1.53	1.24	2.24	2.09	2.87
U07	.	0.74	1.02	1.20	0.82	0.23	0.38	0.71	0.38	0.95	.	0.28
U08	1.37	1.72	2.00	2.42	1.52	1.70	1.78	0.99	0.89	2.09	1.30	1.23
U09	1.80	2.15	2.38	3.37	2.37	1.59	1.28	1.07	0.82	2.13	1.34	0.76
U10	0.87	0.96	1.58	2.11	0.89	0.57	1.40	0.74	0.51	2.00	0.75	0.97
U11	.	0.66	1.17	1.54	0.61	0.44	0.86	0.63	0.38	1.51	0.49	0.71
U12	.	0.77	0.74	1.17	0.34	1.00	1.19	0.47	0.40	1.30	0.42	0.25
U13	.	1.48	1.79	2.46	1.27	1.38	1.69	0.91	0.77	1.99	0.98	1.01
U14	.	1.01	0.73	1.50	0.46	1.06	1.65	0.55	0.50	1.49	0.58	0.26
U15	.	0.34	0.33	0.49	0.12	0.69	0.25	0.29	0.27	0.48	0.15	0.13
U16	.	0.72	1.92	1.09	1.86	0.70	0.71	1.30	0.67	1.20	.	0.42
U17	.	1.81	2.33	3.65	1.27	1.29	2.57	0.83	0.90	3.02	1.64	2.50
U18	.	1.91	3.13	4.53	2.47	0.89	1.77	1.08	0.87	3.05	.	2.55
U19	1.73	1.32	4.02	2.95	2.38	0.51	1.13	1.52	0.87	1.94	1.82	2.45
U20	2.33	1.51	3.88	2.78	2.82	0.94	1.45	1.51	1.09	2.12	2.32	1.95
U21	2.00	1.61	3.52	3.14	3.07	0.79	1.23	1.45	1.05	2.52	1.63	2.07
U22	2.00	1.31	3.20	2.78	2.02	0.56	1.11	1.36	0.93	1.84	1.32	1.92
U23	1.63	2.21	2.47	3.61	2.34	1.92	1.79	1.25	1.14	2.65	1.92	1.62

TABLE D.2 CONTINUED

ISDME CODE	VALIDATION SITE	MODELS										
		TVA	AES	SIMPMOD	STATMOD	MOE	SERTAD	RTM-II	RIVAD	ASTRAP	UMACID	RELMAP
U24	1.23	2.42	3.02	4.96	2.36	1.48	3.11	1.16	1.10	3.99	2.25	2.36
U25	.	0.58	0.51	0.81	0.27	0.87	0.61	0.37	0.30	0.81	0.48	0.12
U26	1.52	1.66	3.76	4.34	2.42	0.62	1.61	1.27	0.87	2.24	1.91	2.69
U40	.	0.48	0.54	1.11	0.42	0.35	0.64	0.52	0.30	1.14	0.39	0.52
U41	.	1.40	1.51	3.20	1.61	0.77	1.54	0.80	0.59	2.61	0.71	1.47
U42	1.20	1.61	2.30	3.38	1.39	1.17	0.32	0.79	0.83	2.83	1.65	1.99
U43	1.70	1.48	1.93	3.88	2.18	0.57	1.26	0.91	0.58	2.37	0.74	1.93
U44	1.77	3.08	6.09	6.25	3.89	1.52	2.66	1.67	1.56	3.20	2.83	3.54
U45	1.77	2.59	2.87	4.54	2.35	2.15	1.97	1.17	1.17	3.23	2.08	2.19
U46	1.43	3.31	4.03	5.57	3.36	2.05	2.12	1.62	1.53	4.12	1.90	3.39
U47	0.90	2.23	2.73	3.67	2.34	1.70	1.69	1.27	1.24	3.20	2.03	1.79
U48	1.27	2.35	2.49	3.50	2.67	1.70	1.64	1.21	1.04	2.80	1.49	1.35
U49	.	1.58	2.00	2.31	1.33	1.50	1.54	0.93	0.85	2.09	1.28	1.10
U50	.	1.01	1.25	1.36	0.81	1.48	0.69	0.80	0.69	1.41	.	0.82
U51	.	0.91	0.83	1.28	0.45	1.16	0.90	0.56	0.51	1.15	.	0.45
U52	0.30	0.30	0.23	0.40	0.13	0.58	0.25	0.28	0.15	0.39	0.19	0.04
U53	.	0.51	0.93	0.74	0.54	0.70	0.30	0.60	0.39	0.74	.	0.28
U54	.	1.23	2.79	1.52	1.67	1.10	0.69	1.21	0.80	1.36	.	0.64
U55	.	1.77	4.58	3.11	3.47	1.24	1.46	1.59	1.29	2.63	2.14	2.06
U56	.	2.08	3.80	3.83	3.70	1.36	1.68	1.51	1.29	2.32	.	2.21
U57	.	1.51	4.10	2.29	2.74	1.17	1.10	1.52	0.97	2.74	.	1.52
U58	.	1.68	4.10	2.93	3.45	1.18	1.59	1.64	1.15	3.10	2.72	2.04
U59	.	1.16	2.69	2.22	2.27	0.41	0.92	1.14	0.69	1.87	.	1.42

TABLE D.3 COMPARISON OF THE OBSERVATIONS AND PREDICTIONS OF **SPRING 1980** SULFUR WET DEPOSITION (KG S/HA) AT THE ISDME EVALUATION SITES.

		MODELS										
ISDME CODE	VALIDATION SITE	TVA	AES	SIMPMOD	STATMOD	MOE	SERTAD	RTM-II	RIVAD	ASTRAP	UMACID	RELMAP
C01	0.73	0.34	0.16	0.27	0.11	0.48	0.56	0.32	0.51	0.35	0.04	0.17
C03	2.83	1.66	2.07	1.66	0.94	1.98	2.15	1.32	1.29	2.71	1.37	3.00
C04	3.38	3.10	1.77	2.76	1.89	3.69	3.36	1.32	1.83	4.18	1.98	3.96
C05	1.32	0.85	0.82	0.95	0.48	0.73	1.03	1.13	0.84	1.85	0.52	2.15
C11	.	0.26	0.06	.	.	0.25	0.57	.	0.16	0.21	.	.
C12	1.63	0.32	0.12	.	.	0.42	0.67	0.39	0.35	0.30	.	.
C13	.	0.52	0.13	0.43	0.19	0.65	0.99	0.49	0.42	0.68	.	0.30
C14	3.40	2.17	2.20	2.00	1.09	2.54	2.25	1.17	1.40	3.05	1.78	3.07
C15	3.93	3.06	2.60	2.86	2.79	3.27	3.10	1.49	1.66	3.81	2.13	3.76
C16	2.93	2.25	2.22	2.06	1.12	2.43	2.15	1.13	1.55	3.65	1.90	2.98
C17	3.13	1.91	2.80	1.85	0.99	1.94	2.23	1.29	1.56	3.50	2.04	2.93
C18	0.43	0.16	0.02	0.15	0.05	0.07	0.56	.	0.00	0.11	.	.
C19	1.03	0.33	0.13	0.32	0.17	0.33	1.00	.	0.33	0.51	.	0.28
C20	1.37	0.73	0.86	0.71	0.39	0.86	1.40	0.82	0.78	1.38	0.13	1.31
C22	1.03	0.21	0.10	0.23	0.20	0.20	1.33	.	0.24	0.34	.	0.16
C23	2.23	0.70	0.71	0.70	0.49	0.93	1.46	1.02	0.85	1.58	0.29	1.00
C24	1.97	0.82	1.00	0.86	0.47	1.07	1.39	1.04	0.96	1.97	0.52	1.53
C26	1.60	0.31	0.16	0.34	0.34	0.30	1.46	.	0.34	0.72	.	.
C33	.	1.48	.	1.45	0.77	1.70	.	.	.	2.94	.	2.10
C34	.	0.72	0.80	0.70	0.37	0.90	1.30	0.87	0.84	1.64	.	1.08
U01	1.67	2.68	3.11	2.84	1.78	2.59	2.41	1.60	1.75	4.74	1.40	4.24
U02	2.17	2.21	1.87	2.15	2.05	1.33	2.41	1.28	1.06	1.77	1.61	2.24
U03	2.73	2.64	2.12	2.21	1.96	2.58	2.76	1.41	1.68	2.91	1.34	3.85
U04	1.13	3.16	2.94	2.94	1.77	2.32	3.05	1.51	1.40	2.80	1.99	4.17
U05	1.40	1.68	2.21	1.62	0.86	1.66	1.88	1.44	1.49	3.29	0.75	2.92
U06	2.40	2.79	2.41	2.82	1.60	2.24	2.40	1.33	1.31	2.86	1.46	5.88
U07	.	0.99	1.08	0.93	0.83	0.31	1.90	1.25	0.64	0.88	.	0.64
U08	3.98	2.60	2.12	2.13	1.80	2.35	2.83	1.46	1.48	2.58	1.19	2.74
U09	3.17	2.99	2.05	2.87	2.71	1.95	2.40	1.33	1.53	2.03	1.06	2.03
U10	2.07	1.54	1.89	1.56	0.89	1.50	2.09	1.39	1.40	3.17	0.51	3.71
U11	.	1.11	1.11	1.15	0.68	1.23	1.76	1.14	1.18	2.50	0.71	2.27
U12	2.07	1.27	1.49	0.98	0.75	2.19	1.93	1.04	1.12	1.72	0.80	1.28
U13	.	2.36	2.16	1.94	1.54	2.54	2.84	1.36	1.58	2.65	1.35	3.24
U14	2.77	1.65	1.75	1.26	1.04	2.28	2.46	1.16	1.27	2.02	0.87	1.50
U15	2.20	0.51	0.64	0.48	0.37	0.86	1.01	0.78	0.83	0.62	0.27	0.26
U16	.	0.99	1.39	0.96	1.50	0.54	2.75	1.35	0.71	0.89	.	0.66
U17	2.20	2.81	2.58	2.67	1.44	2.85	2.52	1.42	1.80	4.28	1.75	4.31
U18	.	2.89	3.12	3.23	2.46	2.49	2.64	1.55	1.52	4.93	.	5.22
U19	2.07	1.82	1.54	1.99	1.30	0.93	1.80	0.91	0.84	2.00	0.78	2.53
U20	3.43	2.08	1.79	2.12	1.59	1.00	2.21	1.03	1.01	1.92	1.25	2.60
U21	3.03	2.24	1.95	2.26	1.67	1.09	2.07	1.04	0.87	2.26	0.62	2.76
U22	1.20	1.83	1.31	1.93	1.19	0.87	1.78	0.97	0.78	1.76	0.42	1.88
U23	3.63	3.28	3.13	2.54	2.35	3.16	2.94	1.80	1.81	3.54	2.42	4.33

TABLE D.3 CONTINUED

		MODELS										
ISOME CODE	VALIDATION SITE	TVA	AES	SIMPMOD	STATMOD	MOE	SERTAD	RTM-II	RIVAD	ASTRAP	UMACID	RELMAP
U24	3.19	3.84	3.61	3.76	2.57	3.62	3.90	1.68	1.88	6.06	2.41	4.83
U25	.	0.93	0.66	0.74	0.63	1.51	1.46	0.77	0.84	1.03	0.65	0.54
U26	3.23	2.37	2.81	2.73	1.90	1.63	2.44	1.37	1.20	3.36	2.82	4.92
U40	1.43	0.83	0.91	0.86	0.48	1.03	1.47	1.01	0.96	1.90	0.39	1.71
U41	3.27	2.20	2.68	2.36	1.83	1.99	2.26	1.51	1.59	4.01	1.23	4.36
U42	2.97	2.52	2.29	2.48	1.54	2.46	1.46	1.47	1.77	4.20	1.28	4.53
U43	2.57	2.25	3.24	2.72	2.22	1.74	2.33	1.44	1.29	4.16	1.13	6.19
U44	5.27	4.42	3.62	4.50	3.13	3.46	3.47	1.67	1.78	4.56	3.22	12.53
U45	2.83	3.80	3.26	3.27	3.03	3.84	3.41	1.91	2.00	4.13	2.23	5.91
U46	4.17	4.85	4.10	3.83	3.66	4.02	3.33	2.07	1.93	5.30	3.07	10.39
U47	3.83	3.46	2.59	2.81	2.46	2.60	2.89	1.58	1.69	3.88	2.44	4.79
U48	3.30	3.30	2.43	2.97	2.89	2.13	2.79	1.56	1.57	3.10	1.74	3.68
U49	3.13	2.34	2.63	2.11	2.23	2.59	2.61	1.44	1.47	3.00	1.40	3.50
U50	.	1.52	2.14	1.40	1.33	1.74	1.74	1.26	1.36	1.60	.	1.78
U51	.	1.40	1.53	1.18	1.08	2.05	1.93	1.10	1.28	1.52	.	1.72
U52	.	0.47	0.23	0.36	0.20	0.67	0.74	0.41	0.56	0.48	0.04	0.25
U53	.	0.68	1.65	0.71	1.03	0.81	1.67	1.14	0.88	0.65	.	0.62
U54	.	1.58	1.68	1.43	1.85	1.32	2.40	0.90	1.08	1.09	.	1.35
U55	.	2.39	2.14	2.57	1.89	1.39	2.18	1.19	1.34	2.42	1.04	3.66
U56	.	2.85	2.61	2.65	1.91	1.92	2.50	1.43	1.49	2.56	.	5.33
U57	.	1.97	1.87	1.96	2.16	1.17	2.59	1.48	0.84	2.04	.	3.29
U58	.	2.28	1.84	2.34	1.71	1.26	2.36	1.23	1.16	2.57	1.68	3.62
U59	.	1.60	1.45	1.65	1.28	0.54	1.86	0.99	0.67	1.56	.	1.98

TABLE D.4 COMPARISON OF THE OBSERVATIONS AND PREDICTIONS OF SUMMER 1980 SULFUR WET DEPOSITION (KG S/HA) AT THE ISDME EVALUATION SITES.

		MODELS										
ISDME CODE	VALIDATION SITE	TVA	AES	SIMPMOD	STATMOD	MOE	SERTAD	RTM-II	RIVAD	ASTRAP	UMACID	RELMAP
C01	0.95	0.31	0.30	0.24	0.29	0.58	1.79	0.89	0.71	0.47	0.27	0.05
C03	3.70	1.97	3.00	1.66	1.25	2.43	3.41	1.63	1.70	3.61	1.63	3.80
C04	3.83	3.81	3.75	2.59	2.15	4.26	5.49	2.04	2.21	5.56	4.19	7.18
C05	1.60	1.05	1.92	1.03	0.68	1.00	1.88	1.36	1.13	2.00	0.82	1.67
C11	.	0.23	0.16	.	.	0.26	1.72	.	0.15	0.29	.	.
C12	.	0.29	0.16	.	.	0.41	1.84	0.83	0.57	0.38	.	.
C13	0.90	0.54	0.18	0.34	0.20	0.78	1.77	0.73	0.51	0.80	.	0.03
C14	3.43	2.47	2.69	1.62	1.15	2.93	3.63	1.58	1.65	3.43	2.67	3.08
C15	6.43	3.62	3.59	3.33	4.00	4.11	4.89	2.12	2.15	5.63	3.53	5.08
C16	4.17	2.69	3.24	1.82	1.11	2.85	3.91	1.56	1.92	4.52	2.73	3.38
C17	4.77	2.36	3.31	1.69	0.96	2.39	3.52	1.48	1.82	4.44	3.07	3.14
C18	0.30	0.16	0.03	0.11	0.05	0.12	0.86	.	0.00	0.18	.	.
C19	0.90	0.38	0.30	0.26	0.21	0.56	1.93	.	0.30	0.64	.	0.10
C20	1.73	0.90	0.66	0.68	0.63	1.20	2.86	1.16	0.91	1.81	0.37	1.47
C22	0.97	0.25	0.12	0.22	0.16	0.21	1.66	.	0.26	0.42	.	0.07
C23	3.70	0.85	0.96	0.69	0.67	1.22	2.68	1.24	1.02	2.01	0.66	0.61
C24	4.17	1.01	1.54	0.90	0.73	1.46	2.57	1.38	1.21	2.27	0.76	1.23
C26	.	0.40	0.40	0.40	0.45	0.38	2.49	.	0.50	0.94	.	.
C33	.	1.78	.	1.37	0.99	1.99	.	.	.	3.68	.	2.14
C34	.	0.87	0.69	0.68	0.64	1.17	2.73	1.16	0.95	2.16	.	0.73
U01	2.10	3.51	2.93	3.10	1.56	2.64	4.04	1.60	2.04	5.48	1.65	6.50
U02	.	2.56	2.36	2.28	1.68	0.98	2.15	1.13	0.95	1.82	1.50	3.84
U03	3.47	3.21	3.44	2.22	2.32	3.27	4.32	2.11	2.11	4.15	2.38	5.08
U04	2.80	3.57	4.63	2.94	2.69	2.18	4.14	2.20	1.87	3.85	4.78	6.32
U05	3.62	2.00	3.91	1.57	1.01	2.00	3.13	1.80	1.78	4.07	2.52	2.53
U06	2.87	3.22	3.65	2.92	1.51	1.85	2.87	1.89	1.74	3.29	1.86	9.15
U07	.	1.04	1.30	0.97	1.10	0.34	2.81	1.89	0.61	1.08	.	1.88
U08	3.03	3.35	2.70	2.29	2.43	2.88	3.52	1.82	1.75	3.47	1.77	4.11
U09	2.30	3.66	2.84	3.00	3.48	1.92	2.71	1.57	1.10	2.75	1.28	3.08
U10	2.80	1.84	3.66	1.51	1.08	1.80	2.92	2.03	1.76	3.76	1.50	3.66
U11	.	1.33	2.42	1.15	1.01	1.59	3.12	1.86	1.55	2.98	1.35	1.63
U12	1.73	1.44	1.16	0.90	0.76	2.37	3.31	1.16	1.42	1.89	0.56	0.91
U13	4.43	2.88	2.94	2.13	1.95	3.50	4.71	1.89	2.18	3.43	1.40	5.72
U14	.	1.98	1.89	1.28	1.21	2.87	4.24	1.53	1.76	2.33	1.31	1.91
U15	0.83	0.46	0.36	0.34	0.35	0.85	1.88	0.84	1.39	0.59	0.27	0.16
U16	.	1.07	1.57	1.16	0.79	0.39	2.07	1.21	0.69	1.06	.	1.61
U17	1.87	3.60	3.32	2.67	1.51	3.21	4.78	1.71	2.25	6.06	2.54	6.76
U18	.	3.70	3.03	3.45	2.22	2.35	3.10	1.57	1.84	4.91	.	6.81
U19	1.47	2.09	2.47	1.71	0.90	0.79	2.21	1.32	1.10	1.74	1.25	4.44
U20	2.10	2.51	3.40	2.11	1.40	0.94	3.33	1.53	1.34	2.07	2.84	6.51
U21	3.57	2.60	3.20	2.10	1.52	1.03	3.42	1.53	1.27	2.45	1.95	4.49
U22	1.93	2.14	2.27	1.62	0.91	0.76	2.31	1.29	1.10	1.70	0.76	3.46
U23	7.87	3.97	3.80	2.63	2.61	3.59	4.47	2.29	2.35	5.75	2.73	6.30

TABLE D.4 CONTINUED

ISDME CODE	VALIDATION SITE	MODELS										
		TVA	AES	SIMPMOD	STATMOD	MOE	SERTAD	RTM-II	RIVAD	ASTRAP	UMACID	RELMAP
U24	4.16	5.22	4.62	4.22	2.83	4.01	5.30	2.32	2.51	9.25	3.82	8.43
U25	2.00	0.99	1.14	0.68	0.89	1.77	3.08	1.26	1.26	1.34	0.77	0.52
U26	1.47	2.92	3.13	3.07	1.87	1.45	2.66	1.42	1.42	3.09	0.82	5.57
U40	.	1.00	1.12	0.87	0.70	1.38	2.66	1.42	1.24	2.29	0.68	1.26
U41	2.73	2.80	2.92	2.48	1.52	2.10	2.76	1.43	1.67	4.44	1.14	4.36
U42	3.97	3.21	2.97	2.51	1.19	2.80	2.49	1.54	1.98	5.51	2.37	5.81
U43	1.13	2.92	2.96	3.11	2.32	1.85	2.26	1.34	1.40	4.07	1.33	6.07
U44	5.43	5.21	6.40	4.82	3.80	3.66	4.30	2.50	2.49	7.32	4.08	17.82
U45	5.20	4.52	3.64	2.90	2.85	4.41	5.31	2.45	2.41	6.04	2.96	6.90
U46	6.50	6.09	4.86	3.55	3.49	4.48	5.31	2.85	2.73	9.49	4.42	13.68
U47	.	4.39	3.88	2.97	2.93	3.14	4.06	2.43	2.60	6.96	3.14	7.66
U48	3.00	4.16	3.33	2.95	2.55	2.34	3.09	1.62	1.63	3.35	1.66	5.32
U49	5.53	2.72	2.68	1.97	2.67	3.26	4.49	1.85	1.81	3.41	1.66	4.62
U50	.	1.93	2.20	1.99	2.93	1.99	2.15	1.31	1.43	1.76	.	2.65
U51	.	1.52	1.77	1.08	1.43	2.40	3.78	1.62	1.72	1.91	.	1.73
U52	1.07	0.44	0.54	0.31	0.42	0.87	2.12	0.96	0.96	0.65	0.19	0.16
U53	.	0.75	0.75	0.75	0.50	0.38	1.01	0.63	0.50	0.30	.	0.67
U54	.	1.70	2.12	1.82	1.61	0.84	1.89	0.76	0.68	0.83	.	1.53
U55	2.17	2.88	2.55	2.63	1.99	1.25	2.59	1.42	1.43	2.56	1.59	6.33
U56	.	3.43	4.40	2.63	2.01	1.90	3.11	1.93	1.89	3.75	.	8.90
U57	.	2.30	2.39	2.53	1.92	1.01	2.09	1.08	0.97	2.30	.	7.78
U58	.	2.82	2.80	2.22	1.37	1.12	2.95	1.36	1.41	2.41	1.35	6.26
U59	.	1.84	2.49	1.53	1.20	0.57	2.58	1.42	0.95	1.84	.	2.88

TABLE D.5 COMPARISON OF THE OBSERVATIONS AND PREDICTIONS OF AUTUMN 1980 SULFUR WET DEPOSITION (KG S/HA) AT THE ISDME EVALUATION SITES.

ISDME CODE	VALIDATION SITE	MODELS										
		TVA	AES	SIMPMOD	STATMOD	MOE	SERTAD	RTM-II	RIVAD	ASTRAP	UMACID	RELMAP
C01	0.30	0.18	0.18	0.18	0.10	0.47	0.65	0.24	0.17	0.31	0.17	0.02
C03	1.60	0.98	1.54	1.62	1.12	1.45	2.49	0.84	0.65	2.31	1.27	1.30
C04	2.03	1.86	2.02	2.38	1.33	1.93	2.57	0.68	0.69	3.18	1.45	1.07
C05	2.20	0.60	1.32	1.04	0.99	0.51	2.20	1.03	0.48	1.60	0.99	1.47
C11	.	0.14	0.13	.	.	0.33	0.88	.	0.05	0.18	.	.
C12	0.43	0.16	0.22	.	.	0.45	0.89	0.24	0.14	0.28	.	.
C13	0.57	0.26	0.20	0.29	0.14	0.67	1.39	0.47	0.25	0.71	.	0.10
C14	2.27	1.27	1.38	1.49	1.02	1.70	2.63	0.77	0.63	2.48	1.95	1.19
C15	2.17	1.70	1.39	2.62	1.71	1.65	1.89	0.57	0.56	2.79	1.32	0.78
C16	3.00	1.31	2.01	1.79	1.17	1.62	3.08	0.77	0.78	3.10	2.17	1.02
C17	3.17	1.14	2.07	1.73	1.13	1.43	2.87	0.82	0.75	2.87	2.31	1.30
C18	.	0.09	0.09	0.11	0.07	0.13	1.21	.	0.00	0.12	.	.
C19	0.57	0.19	0.24	0.28	0.12	0.45	1.28	.	0.16	0.47	.	0.24
C20	1.43	0.41	0.89	0.67	0.49	0.76	2.06	0.71	0.45	1.26	0.22	0.73
C22	0.83	0.13	0.21	0.24	0.19	0.23	1.83	.	0.16	0.35	.	0.18
C23	1.90	0.44	0.90	0.75	0.55	0.70	1.90	0.78	0.52	1.36	0.63	1.14
C24	2.00	0.52	1.04	0.97	0.75	0.72	2.18	0.90	0.57	1.75	0.77	1.33
C26	.	0.21	0.69	0.44	0.56	0.22	2.52	.	0.19	0.69	.	.
C33	.	0.87	.	1.44	0.98	1.29	.	.	.	2.42	.	1.21
C34	.	0.41	1.21	0.73	0.44	0.78	1.76	0.74	0.51	1.41	.	0.83
U01	0.90	1.65	2.66	2.85	1.79	1.55	2.68	0.79	0.78	2.95	1.71	1.79
U02	.	1.47	1.58	2.32	1.97	0.70	1.75	0.62	0.47	1.35	1.31	1.25
U03	1.20	1.45	1.54	2.00	1.33	1.20	1.75	0.57	0.54	2.04	1.03	0.71
U04	0.57	2.07	2.43	2.78	1.63	1.11	1.81	0.72	0.58	1.72	1.65	2.05
U05	1.10	1.01	2.68	1.61	1.06	1.29	2.61	0.91	0.81	2.61	1.23	1.32
U06	0.67	1.86	2.35	2.88	1.56	1.00	1.69	0.73	0.59	1.85	1.07	2.18
U07	.	0.65	0.65	1.02	0.60	0.14	1.04	0.51	0.25	0.80	.	0.33
U08	0.67	1.38	0.90	1.66	1.09	1.04	1.73	0.50	0.41	1.54	0.96	0.48
U09	.	1.75	1.38	2.52	1.89	0.79	1.39	0.61	0.45	1.34	0.97	0.59
U10	1.37	0.94	2.38	1.58	1.18	1.10	2.76	0.99	0.85	2.46	0.91	1.70
U11	0.93	0.67	1.90	1.25	1.01	0.97	2.54	0.97	0.76	1.95	0.97	1.56
U12	.	0.64	0.75	0.67	0.34	1.08	1.58	0.53	0.40	1.38	0.34	0.38
U13	1.57	1.26	1.60	1.66	0.96	1.19	1.84	0.54	0.49	1.93	1.17	0.60
U14	1.17	0.84	0.91	0.89	0.47	1.05	1.63	0.48	0.40	1.52	0.57	0.41
U15	.	0.25	0.10	0.29	0.10	0.35	0.50	0.11	0.14	0.35	0.03	0.03
U16	.	0.66	0.80	1.02	1.12	0.30	1.88	0.60	0.35	0.68	.	0.53
U17	0.77	1.72	2.53	2.60	1.44	1.69	2.70	0.82	0.81	3.12	2.12	1.56
U18	.	1.87	2.41	3.34	2.35	1.29	2.03	0.77	0.67	2.89	.	2.12
U19	0.93	1.25	2.35	1.89	1.52	0.44	1.71	0.92	0.42	1.50	0.83	2.00
U20	.	1.36	1.86	1.96	1.26	0.52	1.49	0.59	0.46	1.08	1.15	1.23
U21	1.43	1.42	1.94	2.06	1.29	0.53	1.40	0.66	0.47	1.42	1.17	1.52
U22	1.52	1.23	1.72	1.83	1.16	0.38	1.40	0.71	0.42	1.29	0.88	1.52
U23	1.30	1.83	1.57	2.44	1.62	1.56	2.01	0.64	0.72	2.44	1.63	1.08

TABLE D.5 CONTINUED

		MODELS										
ISDME CODE	VALIDATION SITE	TVA	AES	SIMPMOD	STATMOD	MOE	SERTAD	RTM-II	RIVAD	ASTRAP	UMACID	RELMAP
U24	1.13	2.35	2.59	3.84	2.26	1.89	2.26	0.82	0.80	3.85	2.22	1.84
U25	0.73	0.44	0.45	0.47	0.21	0.87	0.86	0.31	0.24	0.78	0.34	0.10
U26	0.90	1.64	2.43	3.04	2.18	0.75	2.00	0.78	0.58	2.08	0.67	2.10
U40	.	0.51	1.35	0.96	0.61	0.82	2.00	0.80	0.62	1.57	0.75	1.34
U41	1.27	1.36	2.38	2.36	1.60	1.23	2.53	0.85	0.78	2.85	1.13	1.49
U42	1.42	1.55	2.71	2.46	1.52	1.55	2.52	0.82	0.82	3.04	1.88	1.62
U43	0.83	1.47	2.75	3.04	2.51	0.99	2.13	0.88	0.66	2.58	1.07	2.19
U44	2.03	2.84	2.82	4.50	2.86	1.66	1.92	0.74	0.77	2.57	2.76	2.82
U45	1.23	2.11	1.90	2.64	1.59	2.00	2.07	0.71	0.73	2.88	1.53	1.34
U46	1.37	2.82	2.01	3.27	2.34	2.01	2.22	0.77	0.72	3.68	1.89	2.34
U47	0.93	1.94	1.54	2.57	1.66	1.18	1.75	0.61	0.59	2.63	1.26	1.05
U48	1.07	1.92	1.48	2.63	2.14	0.98	1.74	0.66	0.50	2.16	1.37	1.11
U49	.	1.20	1.03	1.37	0.98	1.09	1.50	0.42	0.41	1.73	1.26	0.48
U50	.	0.75	0.60	0.93	0.76	0.71	1.25	0.44	0.33	0.82	.	0.30
U51	.	0.65	0.48	0.71	0.38	0.81	1.08	0.36	0.28	0.95	.	0.25
U52	.	0.23	0.23	0.24	0.10	0.54	0.60	0.19	0.15	0.42	0.16	0.03
U53	.	0.40	0.72	0.53	0.62	0.34	1.31	0.54	0.34	0.43	.	0.30
U54	.	0.95	1.42	1.40	1.63	0.58	2.00	0.60	0.38	1.01	.	1.01
U55	1.80	1.55	1.93	2.37	1.76	0.73	1.63	0.65	0.54	1.44	0.94	1.53
U56	.	1.82	1.78	2.53	1.87	0.96	1.69	0.66	0.57	1.61	.	1.62
U57	.	1.28	1.51	2.05	1.73	0.63	1.70	0.57	0.44	1.53	.	1.30
U58	.	1.48	1.81	2.18	1.49	0.65	1.59	0.63	0.49	1.42	1.06	1.52
U59	.	1.06	1.34	1.57	0.83	0.28	1.12	0.53	0.30	1.07	.	0.96

TABLE D.6 STANDARDIZED RESIDUALS OF ANNUAL 1980 SULFUR WET DEPOSITION AT THE ISDME EVALUATION SITES.

STANDARDIZED RESIDUAL = (VALIDATION-MODEL)/VALIDATION VALUE

ANNUAL

ISDME CODE	VALIDATION SITE	MODELS										
		TVA	AES	SIMPMOD	STATMOD	MOE	SERTAD	RTM-II	RIVAD	ASTRAP	UMACID	RELMAP
C01	2.05	0.43	0.61	0.53	0.74	0.01	-0.60	0.17	0.26	0.32	0.70	0.87
C03	9.00	0.29	0.11	0.25	0.56	0.22	-0.13	0.51	0.53	-0.16	0.41	-0.00
C04	11.67	-0.09	0.20	0.04	0.35	0.01	-0.01	0.58	0.52	-0.38	0.19	-0.20
C05	5.83	0.41	0.17	0.28	0.57	0.59	-0.05	0.30	0.53	-0.13	0.52	-0.05
C11	.	.	.	.	.	.	.	.	.	.	.	.
C12	.	.	.	.	.	.	.	.	.	.	.	.
C13	.	.	.	.	.	.	.	.	.	.	.	.
C14	11.03	0.24	0.30	0.34	0.62	0.23	0.01	0.63	0.62	-0.01	0.28	0.22
C15	15.77	0.24	0.37	0.21	0.29	0.31	0.34	0.67	0.67	0.05	0.44	0.30
C16	12.20	0.28	0.28	0.34	0.65	0.34	-0.00	0.67	0.60	-0.13	0.28	0.27
C17	13.50	0.45	0.27	0.45	0.72	0.50	0.17	0.68	0.65	0.02	0.34	0.36
C18	.	.	.	.	.	.	.	.	.	.	.	.
C19	2.77	0.57	0.71	0.58	0.80	0.39	-0.87	.	0.67	0.27	.	0.74
C20	5.13	0.48	0.42	0.47	0.67	0.31	-0.49	0.38	0.51	-0.05	0.85	0.26
C22	3.53	0.79	0.85	0.74	0.82	0.80	-0.73	.	0.79	0.62	.	0.85
C23	9.20	0.71	0.68	0.68	0.78	0.66	0.20	0.62	0.71	0.36	0.80	0.66
C24	9.13	0.65	0.55	0.59	0.75	0.62	0.16	0.58	0.67	0.21	0.74	0.49
C26	.	.	.	.	.	.	.	.	.	.	.	.
C33	.	.	.	.	.	.	.	.	.	.	.	.
C34	.	.	.	.	.	.	.	.	.	.	.	.
U01	5.57	-0.93	-1.02	-1.20	-0.16	-0.42	-0.97	0.12	0.02	-1.89	-0.10	-1.69
U02	.	.	.	.	.	.	.	.	.	.	.	.
U03	8.63	-0.18	-0.09	-0.07	0.15	0.00	-0.26	0.40	0.39	-0.32	0.30	-0.27
U04	5.33	-1.41	-1.87	-1.38	-0.57	-0.31	-1.08	-0.14	0.04	-0.99	-1.19	-1.87
U05	7.47	0.13	-0.40	0.09	0.51	0.23	-0.34	0.34	0.38	-0.62	0.30	-0.05
U06	7.03	-0.62	-0.91	-0.79	0.04	0.12	-0.36	0.22	0.31	-0.46	0.09	-1.85
U07	.	.	.	.	.	.	.	.	.	.	.	.
U08	9.93	-0.03	0.22	0.17	0.35	0.20	0.10	0.52	0.54	0.03	0.47	0.14
U09	.	.	.	.	.	.	.	.	.	.	.	.
U10	7.13	0.17	-0.33	0.08	0.46	0.30	-0.27	0.28	0.37	-0.60	0.48	-0.41
U11	.	.	.	.	.	.	.	.	.	.	.	.
U12	.	.	.	.	.	.	.	.	.	.	.	.
U13	.	.	.	.	.	.	.	.	.	.	.	.
U14	.	.	.	.	.	.	.	.	.	.	.	.
U15	.	.	.	.	.	.	.	.	.	.	.	.
U16	.	.	.	.	.	.	.	.	.	.	.	.
U17	.	.	.	.	.	.	.	.	.	.	.	.
U18	.	.	.	.	.	.	.	.	.	.	.	.
U19	6.23	-0.19	-0.67	-0.32	0.14	0.57	-0.22	0.25	0.48	-0.15	0.25	-0.83
U20	.	.	.	.	.	.	.	.	.	.	.	.

TABLE D.6 CONTINUED

STANDARDIZED RESIDUAL = (VALIDATION-MODEL)/VALIDATION VALUE

ANNUAL

ISDME CODE	VALIDATION SITE	TVA	AES	SIMPMOD	STATMOD	MOE	SERTAD	RTM-II	RIVAD	ASTRAP	UMACID	RELMAP
		MODELS										
U21	10.07	0.11	-0.05	0.08	0.31	0.66	0.16	0.54	0.64	0.14	0.45	-0.08
U22	6.23	-0.19	-0.36	-0.27	0.24	0.59	-0.16	0.31	0.48	-0.06	0.45	-0.41
U23	14.40	0.11	0.24	0.24	0.40	0.29	0.24	0.59	0.58	0.00	0.39	0.07
U24	9.25	-0.69	-0.50	-0.76	-0.10	-0.19	-0.35	0.35	0.32	-1.50	-0.16	-0.89
U25	.	.	.	.	.	.	.	.	.	.	.	.
U26	6.93	-0.42	-0.75	-0.83	-0.17	0.36	-0.19	0.30	0.41	-0.55	0.10	-1.20
U40	.	.	.	.	.	.	.	.	.	.	.	.
U41	.	.	.	.	.	.	.	.	.	.	.	.
U42	8.33	-0.21	-0.23	-0.26	0.35	0.04	0.06	0.45	0.35	-0.87	0.14	-0.67
U43	6.23	-0.48	-0.75	-0.99	-0.47	0.17	-0.31	0.27	0.37	-1.11	0.31	-1.63
U44	14.53	-0.24	-0.30	-0.34	0.05	0.29	0.20	0.55	0.55	-0.21	0.11	-1.53
U45	11.00	-0.35	-0.06	-0.17	0.11	-0.13	-0.03	0.43	0.43	-0.48	0.20	-0.49
U46	13.47	-0.46	-0.11	-0.16	0.01	0.07	0.09	0.46	0.49	-0.68	0.16	-1.21
U47	.	.	.	.	.	.	.	.	.	.	.	.
U48	8.67	-0.55	-0.12	-0.36	-0.15	0.17	-0.06	0.41	0.45	-0.32	0.28	-0.32
U49	.	.	.	.	.	.	.	.	.	.	.	.
U50	.	.	.	.	.	.	.	.	.	.	.	.
U51	.	.	.	.	.	.	.	.	.	.	.	.
U52	.	.	.	.	.	.	.	.	.	.	.	.
U53	.	.	.	.	.	.	.	.	.	.	.	.
U54	.	.	.	.	.	.	.	.	.	.	.	.
U55	.	.	.	.	.	.	.	.	.	.	.	.
U56	.	.	.	.	.	.	.	.	.	.	.	.
U57	.	.	.	.	.	.	.	.	.	.	.	.
U58	.	.	.	.	.	.	.	.	.	.	.	.
U59	.	.	.	.	.	.	.	.	.	.	.	.

TABLE D.7 STANDARDIZED RESIDUALS OF WINTER 1980 SULFUR WET DEPOSITION AT THE ISDME EVALUATION SITES.

STANDARDIZED RESIDUAL = (VALIDATION-MODEL)/VALIDATION VALUE

WINTER

ISDME CODE	VALIDATION SITE	MODELS										
		TVA	AES	SIMPMOD	STATMOD	MOE	SERTAD	RTM-II	RIVAD	ASTRAP	UMACID	RELMAP
C01	0.32	0.31	0.49	0.05	0.65	-0.58	0.27	0.23	0.59	0.12	0.56	0.91
C03	0.83	-0.25	-0.64	-1.33	0.06	-0.38	-0.54	0.29	0.34	-1.20	-0.21	-0.09
C04	2.38	0.05	0.27	-0.59	0.21	0.32	-0.14	0.62	0.65	-0.31	0.23	0.25
C05	1.23	0.57	0.37	-0.02	0.55	0.85	0.64	0.55	0.77	0.10	0.63	0.34
C11	.	.	.	.	.	.	.	.	.	.	.	.
C12	.	.	.	.	.	.	.	.	.	.	.	.
C13	0.57	0.44	0.49	0.29	0.82	-0.46	0.29	0.42	0.36	-0.13	.	0.21
C14	1.93	0.27	0.27	-0.23	0.53	0.33	0.22	0.69	0.71	-0.12	0.25	0.35
C15	3.20	0.35	0.26	-0.22	0.28	0.44	0.46	0.70	0.73	0.14	0.43	0.55
C16	2.23	0.35	0.39	-0.14	0.63	0.47	0.10	0.73	0.70	-0.14	0.13	0.29
C17	2.57	0.54	0.36	0.09	0.69	0.60	0.34	0.74	0.76	0.08	0.42	0.49
C18	.	.	.	.	.	.	.	.	.	.	.	.
C19	0.27	0.29	0.55	-0.31	0.59	-0.35	-0.13	.	0.55	-0.58	.	0.62
C20	0.60	0.33	0.10	-0.23	0.52	-0.22	-0.15	0.15	0.40	-0.63	0.92	0.50
C22	0.67	0.82	0.85	0.58	0.74	0.91	0.67	.	0.89	0.64	.	0.82
C23	1.37	0.70	0.74	0.36	0.72	0.78	0.66	0.65	0.81	0.33	0.80	0.71
C24	1.00	0.52	0.50	-0.13	0.53	0.74	0.47	0.46	0.71	-0.18	0.68	0.44
C26	.	.	.	.	.	.	.	.	.	.	.	.
C33	.	.	.	.	.	.	.	.	.	.	.	.
C34	.	.	.	.	.	.	.	.	.	.	.	.
U01	0.90	-0.86	-1.80	-3.18	-0.69	-0.24	-1.42	-0.01	-0.00	-2.24	-0.52	-1.72
U02	2.50	0.30	-0.48	-0.04	-0.33	0.50	0.44	0.40	0.62	0.09	0.24	0.49
U03	1.20	-0.43	-0.95	-1.47	-0.62	-0.29	-0.30	0.09	0.22	-0.92	-0.08	-0.09
U04	0.80	-1.91	-5.66	-4.67	-2.27	-0.71	-1.99	-1.04	-0.59	-1.81	-3.04	-2.43
U05	0.67	-0.59	-1.52	-2.24	-0.29	-0.20	-1.13	-0.13	0.13	-2.21	-0.07	-0.57
U06	1.33	-0.53	-2.79	-2.37	-0.91	0.19	-0.59	-0.15	0.07	-0.68	-0.57	-1.15
U07	.	.	.	.	.	.	.	.	.	.	.	.
U08	1.37	-0.26	-0.46	-0.77	-0.11	-0.24	-0.30	0.27	0.35	-0.53	0.05	0.10
U09	1.80	-0.19	-0.32	-0.87	-0.32	0.12	0.29	0.41	0.54	-0.18	0.26	0.58
U10	0.87	-0.11	-0.82	-1.43	-0.03	0.34	-0.62	0.14	0.41	-1.31	0.13	-0.12
U11	.	.	.	.	.	.	.	.	.	.	.	.
U12	.	.	.	.	.	.	.	.	.	.	.	.
U13	.	.	.	.	.	.	.	.	.	.	.	.
U14	.	.	.	.	.	.	.	.	.	.	.	.
U15	.	.	.	.	.	.	.	.	.	.	.	.
U16	.	.	.	.	.	.	.	.	.	.	.	.
U17	.	.	.	.	.	.	.	.	.	.	.	.
U18	.	.	.	.	.	.	.	.	.	.	.	.
U19	1.73	0.24	-1.32	-0.70	-0.37	0.71	0.35	0.12	0.50	-0.12	-0.05	-0.41
U20	2.33	0.35	-0.66	-0.19	-0.21	0.60	0.38	0.35	0.53	0.09	0.01	0.16

TABLE D.7 CONTINUED

STANDARDIZED RESIDUAL = (VALIDATION-MODEL)/VALIDATION VALUE

WINTER

ISDME CODE	VALIDATION SITE	MODELS										
		TVA	AES	SIMPMOD	STATMOD	MOE	SERTAD	RTM-II	RIVAD	ASTRAP	UMACID	RELMAP
U21	2.00	0.19	-0.76	-0.57	-0.54	0.60	0.38	0.27	0.48	-0.26	0.18	-0.04
U22	2.00	0.35	-0.60	-0.39	-0.01	0.72	0.44	0.32	0.53	0.08	0.34	0.04
U23	1.63	-0.35	-0.51	-1.21	-0.43	-0.18	-0.10	0.23	0.30	-0.62	-0.18	0.01
U24	1.23	-0.96	-1.45	-3.02	-0.91	-0.20	-1.52	0.06	0.11	-2.24	-0.82	-0.91
U25	.	.	.	.	.	.	.	.	.	.	.	.
U26	1.52	-0.[illegible]	-1.48	-1.86	-0.60	0.59	-0.06	0.16	0.43	-0.48	-0.26	-0.77
U40	.	.	.	.	.	.	.	.	.	.	.	.
U41	.	.	.	.	.	.	.	.	.	.	.	.
U42	1.20	-0.34	-0.92	-1.82	-0.16	0.03	0.73	0.34	0.31	-1.36	-0.37	-0.66
U43	1.70	0.13	-0.14	-1.28	-0.28	0.66	0.26	0.47	0.66	-0.39	0.56	-0.14
U44	1.77	-0.74	-2.45	-2.54	-1.20	0.14	-0.51	0.05	0.12	-0.81	-0.60	-1.00
U45	1.77	-0.47	-0.62	-1.57	-0.33	-0.22	-0.12	0.34	0.34	-0.83	-0.18	-0.24
U46	1.43	-1.31	-1.81	-2.89	-1.34	-0.43	-0.48	-0.13	-0.07	-1.87	-0.33	-1.37
U47	0.90	-1.48	-2.03	-3.08	-1.60	-0.89	-0.88	-0.41	-0.38	-2.56	-1.26	-0.99
U48	1.27	-0.86	-0.97	-1.76	-1.11	-0.34	-0.29	0.04	0.18	-1.21	-0.18	-0.07
U49	.	.	.	.	.	.	.	.	.	.	.	.
U50	.	.	.	.	.	.	.	.	.	.	.	.
U51	.	.	.	.	.	.	.	.	.	.	.	.
U52	0.30	0.00	0.23	-0.33	0.57	-0.93	0.17	0.06	0.50	-0.30	0.37	0.87
U53	.	.	.	.	.	.	.	.	.	.	.	.
U54	.	.	.	.	.	.	.	.	.	.	.	.
U55	.	.	.	.	.	.	.	.	.	.	.	.
U56	.	.	.	.	.	.	.	.	.	.	.	.
U57	.	.	.	.	.	.	.	.	.	.	.	.
U58	.	.	.	.	.	.	.	.	.	.	.	.
U59	.	.	.	.	.	.	.	.	.	.	.	.

TABLE D.8 STANDARDIZED RESIDUALS OF SPRING 1980 SULFUR WET DEPOSITION AT THE ISDME EVALUATION SITES.

STANDARDIZED RESIDUAL = (VALIDATION-MODEL)/VALIDATION VALUE

SPRING

ISDME CODE	VALIDATION SITE	MODELS										
		TVA	AES	SIMPMOD	STATMOD	MOE	SERTAD	RTM-II	RIVAD	ASTRAP	UMACID	RELMAP
C01	0.73	0.54	0.78	0.63	0.85	0.35	0.24	0.56	0.30	0.52	0.95	0.77
C03	2.83	0.41	0.27	0.41	0.67	0.30	0.24	0.53	0.54	0.04	0.52	-0.06
C04	3.38	0.08	0.48	0.18	0.44	-0.09	0.01	0.61	0.46	-0.24	0.41	-0.17
C05	1.32	0.35	0.38	0.28	0.64	0.45	0.18	0.14	0.36	-0.41	0.61	-0.63
C11	.	.	.	.	.	.	.	.	.	.	.	.
C12	1.63	0.80	0.93	.	.	0.74	0.59	0.76	0.79	0.82	.	.
C13	.	.	.	.	.	.	.	.	.	.	.	.
C14	3.40	0.36	0.35	0.41	0.68	0.25	0.34	0.66	0.59	0.10	0.48	0.10
C15	3.93	0.22	0.34	0.27	0.29	0.17	0.21	0.62	0.58	0.03	0.46	0.04
C16	2.93	0.23	0.24	0.30	0.62	0.17	0.27	0.61	0.47	-0.24	0.35	-0.02
C17	3.13	0.39	0.11	0.41	0.68	0.38	0.29	0.59	0.50	-0.12	0.35	0.06
C18	0.43	0.63	0.95	0.65	0.88	0.84	-0.29	.	1.00	0.75	.	.
C19	1.03	0.68	0.87	0.69	0.84	0.68	0.03	.	0.68	0.51	.	0.73
C20	1.37	0.47	0.37	0.48	0.71	0.37	-0.02	0.40	0.43	-0.01	0.90	0.04
C22	1.03	0.80	0.90	0.78	0.81	0.81	-0.29	.	0.77	0.67	.	0.85
C23	2.23	0.69	0.68	0.69	0.78	0.58	0.35	0.54	0.62	0.29	0.87	0.55
C24	1.97	0.58	0.49	0.56	0.76	0.46	0.29	0.47	0.51	-0.00	0.74	0.22
C26	1.60	0.81	0.90	0.79	0.79	0.81	0.09	.	0.79	0.55	.	.
C33	.	.	.	.	.	.	.	.	.	.	.	.
C34	.	.	.	.	.	.	.	.	.	.	.	.
U01	1.67	-0.61	-0.87	-0.70	-0.07	-0.55	-0.45	0.04	-0.05	-1.84	0.16	-1.54
U02	2.17	-0.02	0.14	0.01	0.05	0.39	-0.11	0.41	0.51	0.18	0.26	-0.03
U03	2.73	0.03	0.22	0.19	0.28	0.06	-0.01	0.48	0.39	-0.06	0.51	-0.41
U04	1.13	-1.79	-1.59	-1.59	-0.56	-1.05	-1.69	-0.33	-0.24	-1.47	-0.76	-2.68
U05	1.40	-0.20	-0.58	-0.16	0.39	-0.19	-0.34	-0.03	-0.06	-1.35	0.46	-1.09
U06	2.40	-0.16	-0.00	-0.17	0.33	0.07	0.00	0.45	0.45	-0.19	0.39	-1.45
U07	.	.	.	.	.	.	.	.	.	.	.	.
U08	3.98	0.35	0.47	0.47	0.55	0.41	0.29	0.63	0.63	0.35	0.70	0.31
U09	3.17	0.06	0.35	0.09	0.14	0.38	0.24	0.58	0.52	0.36	0.67	0.36
U10	2.07	0.25	0.09	0.25	0.57	0.27	-0.01	0.33	0.32	-0.53	0.75	-0.80
U11	.	.	.	.	.	.	.	.	.	.	.	.
U12	2.07	0.39	0.28	0.53	0.64	-0.06	0.07	0.50	0.46	0.17	0.61	0.38
U13	.	.	.	.	.	.	.	.	.	.	.	.
U14	2.77	0.40	0.37	0.54	0.62	0.18	0.11	0.58	0.54	0.27	0.69	0.46
U15	2.20	0.77	0.71	0.78	0.83	0.61	0.54	0.64	0.62	0.72	0.88	0.88
U16	.	.	.	.	.	.	.	.	.	.	.	.
U17	2.20	-0.28	-0.17	-0.21	0.35	-0.30	-0.15	0.35	0.18	-0.95	0.20	-0.96
U18	.	.	.	.	.	.	.	.	.	.	.	.
U19	2.07	0.12	0.25	0.04	0.37	0.55	0.13	0.56	0.59	0.03	0.62	-0.22
U20	3.43	0.39	0.48	0.38	0.54	0.71	0.36	0.70	0.71	0.44	0.64	0.24

TABLE D.8 CONTINUED

STANDARDIZED RESIDUAL = (VALIDATION-MODEL)/VALIDATION VALUE

SPRING

ISDME CODE	VALIDATION SITE	MODELS										
		TVA	AES	SIMPMOD	STATMOD	MOE	SERTAD	RTM-II	RIVAD	ASTRAP	UMACID	RELMAP
U21	3.03	0.26	0.36	0.25	0.45	0.64	0.32	0.66	0.71	0.25	0.80	0.09
U22	1.20	-0.52	-0.09	-0.61	0.01	0.27	-0.48	0.19	0.35	-0.47	0.65	-0.57
U23	3.63	0.10	0.14	0.30	0.35	0.13	0.19	0.50	0.50	0.03	0.33	-0.19
U24	3.19	-0.20	-0.13	-0.18	0.19	-0.14	-0.22	0.47	0.41	-0.90	0.24	-0.51
U25	.	.	.	.	.	.	.	.	.	.	.	.
U26	3.23	0.27	0.13	0.16	0.41	0.50	0.25	0.58	0.63	-0.04	0.13	-0.52
U40	1.43	0.42	0.37	0.40	0.67	0.28	-0.03	0.30	0.33	-0.33	0.73	-0.19
U41	3.27	0.33	0.18	0.28	0.44	0.39	0.31	0.54	0.51	-0.23	0.62	-0.33
U42	2.97	0.15	0.23	0.16	0.48	0.17	0.51	0.50	0.40	-0.42	0.57	-0.53
U43	2.57	0.12	-0.26	-0.06	0.14	0.32	0.09	0.44	0.50	-0.62	0.56	-1.41
U44	5.27	0.16	0.31	0.15	0.41	0.34	0.34	0.68	0.66	0.13	0.39	-1.38
U45	2.83	-0.34	-0.15	-0.15	-0.07	-0.36	-0.20	0.33	0.29	-0.46	0.21	-1.09
U46	4.17	-0.16	0.02	0.08	0.12	0.04	0.20	0.50	0.54	-0.27	0.26	-1.49
U47	3.83	0.10	0.32	0.27	0.36	0.32	0.25	0.59	0.56	-0.01	0.36	-0.25
U48	3.30	0.00	0.26	0.10	0.12	0.35	0.15	0.53	0.52	0.06	0.47	-0.12
U49	3.13	0.25	0.16	0.33	0.29	0.17	0.17	0.54	0.53	0.04	0.55	-0.12
U50	.	.	.	.	.	.	.	.	.	.	.	.
U51	.	.	.	.	.	.	.	.	.	.	.	.
U52	.	.	.	.	.	.	.	.	.	.	.	.
U53	.	.	.	.	.	.	.	.	.	.	.	.
U54	.	.	.	.	.	.	.	.	.	.	.	.
U55	.	.	.	.	.	.	.	.	.	.	.	.
U56	.	.	.	.	.	.	.	.	.	.	.	.
U57	.	.	.	.	.	.	.	.	.	.	.	.
U58	.	.	.	.	.	.	.	.	.	.	.	.
U59	.	.	.	.	.	.	.	.	.	.	.	.

TABLE D.9 STANDARDIZED RESIDUALS OF **SUMMER 1980** SULFUR WET DEPOSITION AT THE ISDME EVALUATION SITES.

STANDARDIZED RESIDUAL = (VALIDATION-MODEL)/VALIDATION VALUE

SUMMER

ISDME CODE	VALIDATION SITE	MODELS										
		TVA	AES	SIMPMOD	STATMOD	MOE	SERTAD	RTM-II	RIVAD	ASTRAP	UMACID	RELMAP
C01	0.95	0.67	0.68	0.75	0.69	0.39	-0.88	0.06	0.25	0.51	0.72	0.95
C03	3.70	0.47	0.19	0.55	0.66	0.34	0.08	0.56	0.54	0.02	0.56	-0.03
C04	3.83	0.01	0.02	0.32	0.44	-0.11	-0.43	0.47	0.42	-0.45	-0.09	-0.87
C05	1.60	0.34	-0.20	0.36	0.57	0.37	-0.18	0.15	0.29	-0.25	0.49	-0.04
C11	.	.	.	.	.	.	.	.	.	.	.	.
C12	.	.	.	.	.	.	.	.	.	.	.	.
C13	0.90	0.40	0.80	0.62	0.78	0.13	-0.97	0.19	0.43	0.11	.	0.97
C14	3.43	0.28	0.22	0.53	0.67	0.15	-0.06	0.54	0.52	0.00	0.22	0.10
C15	6.43	0.44	0.44	0.48	0.38	0.36	0.24	0.67	0.67	0.12	0.45	0.21
C16	4.17	0.35	0.22	0.56	0.73	0.32	0.06	0.63	0.54	-0.08	0.34	0.19
C17	4.77	0.50	0.31	0.65	0.80	0.50	0.26	0.69	0.62	0.07	0.36	0.34
C18	0.30	0.47	0.90	0.63	0.83	0.60	-1.87	.	1.00	0.40	.	.
C19	0.90	0.58	0.67	0.71	0.77	0.38	-1.14	.	0.67	0.29	.	0.89
C20	1.73	0.48	0.62	0.61	0.64	0.31	-0.65	0.33	0.47	-0.04	0.79	0.15
C22	0.97	0.74	0.88	0.77	0.83	0.78	-0.72	.	0.73	0.57	.	0.93
C23	3.70	0.77	0.74	0.81	0.82	0.67	0.28	0.66	0.72	0.46	0.82	0.84
C24	4.17	0.76	0.63	0.78	0.82	0.65	0.38	0.67	0.71	0.46	0.82	0.70
C26	.	.	.	.	.	.	.	.	.	.	.	.
C33	.	.	.	.	.	.	.	.	.	.	.	.
C34	.	.	.	.	.	.	.	.	.	.	.	.
U01	2.10	-0.67	-0.40	-0.48	0.26	-0.26	-0.92	0.24	0.03	-1.61	0.21	-2.10
U02	.	.	.	.	.	.	.	.	.	.	.	.
U03	3.47	0.07	0.01	0.36	0.33	0.06	-0.25	0.39	0.39	-0.20	0.31	-0.47
U04	2.80	-0.28	-0.65	-0.05	0.04	0.22	-0.48	0.21	0.33	-0.37	-0.71	-1.26
U05	3.62	0.45	-0.08	0.57	0.72	0.45	0.13	0.50	0.51	-0.13	0.30	0.30
U06	2.87	-0.12	-0.27	-0.02	0.47	0.35	-0.00	0.34	0.39	-0.15	0.35	-2.19
U07	.	.	.	.	.	.	.	.	.	.	.	.
U08	3.03	-0.10	0.11	0.25	0.20	0.05	-0.16	0.40	0.42	-0.14	0.42	-0.35
U09	2.30	-0.59	-0.23	-0.30	-0.51	0.17	-0.18	0.32	0.52	-0.20	0.44	-0.34
U10	2.80	0.34	-0.31	0.46	0.61	0.36	-0.04	0.28	0.37	-0.34	0.46	-0.31
U11	.	.	.	.	.	.	.	.	.	.	.	.
U12	1.73	0.17	0.33	0.48	0.56	-0.37	-0.91	0.33	0.18	-0.09	0.68	0.47
U13	4.43	0.35	0.34	0.52	0.56	0.21	-0.06	0.57	0.51	0.23	0.68	-0.29
U14	.	.	.	.	.	.	.	.	.	.	.	.
U15	0.83	0.45	0.57	0.59	0.58	-0.02	-1.26	-0.00	-0.67	0.29	0.68	0.81
U16	.	.	.	.	.	.	.	.	.	.	.	.
U17	1.87	-0.93	-0.78	-0.43	0.19	-0.72	-1.56	0.08	-0.21	-2.25	-0.36	-2.62
U18	.	.	.	.	.	.	.	.	.	.	.	.
U19	1.47	-0.43	-0.68	-0.17	0.39	0.46	-0.51	0.10	0.25	-0.19	0.15	-2.03
U20	2.10	-0.20	-0.62	-0.00	0.33	0.55	-0.59	0.27	0.36	0.01	-0.35	-2.10

TABLE D.9 CONTINUED

STANDARDIZED RESIDUAL = (VALIDATION-MODEL)/VALIDATION VALUE

SUMMER

		MODELS										
ISDME CODE	VALIDATION SITE	TVA	AES	SIMPMOD	STATMOD	MOE	SERTAD	RTM-II	RIVAD	ASTRAP	UMACID	RELMAP
U21	3.57	0.27	0.10	0.41	0.57	0.71	0.04	0.57	0.64	0.31	0.45	-0.26
U22	1.93	-0.11	-0.17	0.16	0.53	0.61	-0.19	0.33	0.43	0.12	0.61	-0.79
U23	7.87	0.50	0.52	0.67	0.67	0.54	0.43	0.71	0.70	0.27	0.65	0.20
U24	4.[illegible]	-0.26	-0.11	-0.02	0.32	0.04	-0.28	0.44	0.40	-1.23	0.08	-1.03
U25	2.00	0.50	0.43	0.66	0.55	[illegible]	-0.54	0.37	0.37	0.33	0.61	0.74
U26	1.47	-0.99	-1.13	-1.09	-0.28	[illegible]	-0.81	0.03	0.03	-1.11	0.44	-2.80
U40	.	.	.	.	.	.	.	.	.	.	.	.
U41	2.73	-0.02	-0.07	0.09	0.44	0.23	-0.01	0.48	0.39	-0.62	0.58	-0.60
U42	3.97	0.19	0.25	0.37	0.70	0.29	0.37	0.61	0.50	-0.39	0.40	-0.46
U43	1.13	-1.58	-1.61	-1.74	-1.05	-0.63	-0.99	-0.18	-0.24	-2.59	-0.17	-4.36
U44	5.43	0.04	-0.18	[illegible]	0.30	0.33	0.21	0.54	0.54	-0.35	0.25	-2.28
U45	5.20	0.13	0.30	0.44	0.45	0.15	-0.02	0.53	0.54	-0.16	0.43	-0.33
U46	6.50	0.06	0.25	0.45	0.46	0.31	0.18	0.56	0.58	-0.46	0.32	-1.10
U47	.	.	.	.	.	.	.	.	.	.	.	.
U48	3.00	-0.39	-0.11	0.02	0.15	0.22	-0.03	0.46	0.46	-0.12	0.45	-0.77
U49	5.53	0.51	0.52	0.64	0.52	0.41	0.19	0.67	0.67	0.38	0.70	0.17
U50	.	.	.	.	.	.	.	.	.	.	.	.
U51	.	.	.	.	.	.	.	.	.	.	.	.
U52	1.07	0.59	0.49	0.71	0.61	0.18	-0.99	0.10	0.10	0.39	0.82	0.85
U53	.	.	.	.	.	.	.	.	.	.	.	.
U54	.	.	.	.	.	.	.	.	.	.	.	.
U55	2.17	-0.33	-0.18	-0.21	0.08	0.42	-0.20	0.34	0.34	-0.18	0.27	-1.92
U56	.	.	.	.	.	.	.	.	.	.	.	.
U57	.	.	.	.	.	.	.	.	.	.	.	.
U58	.	.	.	.	.	.	.	.	.	.	.	.
U59	.	.	.	.	.	.	.	.	.	.	.	.

TABLE D.10 STANDARDIZED RESIDUALS OF **AUTUMN 1980** SULFUR WET DEPOSITION AT THE ISDME EVALUATION SITES.

STANDARDIZED RESIDUAL = (VALIDATION-MODEL)/VALIDATION VALUE

FALL

ISDME CODE	VALIDATION SITE	MODELS TVA	AES	SIMPMOD	STATMOD	MOE	SERTAD	RTM-II	RIVAD	ASTRAP	UMACID	RELMAP
C01	0.30	0.40	0.40	0.40	0.67	-0.57	-1.17	0.21	0.43	-0.03	0.43	0.93
C03	1.60	0.39	0.04	-0.01	0.30	0.09	-0.56	0.47	0.59	-0.44	0.21	0.19
C04	2.03	0.09	0.01	-0.17	0.35	0.05	-0.26	0.67	0.66	-0.56	0.29	0.47
C05	2.20	0.73	0.40	0.53	0.55	0.77	0.00	0.53	0.78	0.27	0.55	0.33
C11	.	.	.	.	.	.	.	.	.	.	.	.
C12	0.43	0.63	0.49	.	.	-0.04	-1.05	0.45	0.68	0.35	.	.
C13	0.57	0.54	0.65	0.49	0.75	-0.18	-1.45	0.18	0.56	-0.25	.	0.82
C14	2.27	0.44	0.39	0.34	0.55	0.25	-0.16	0.66	0.72	-0.09	0.14	0.48
C15	2.17	0.22	0.36	-0.21	0.21	0.24	0.13	0.74	0.74	-0.29	0.39	0.64
C16	3.00	0.56	0.33	0.40	0.61	0.46	-0.03	0.74	0.74	-0.03	0.28	0.66
C17	3.17	0.64	0.35	0.45	0.64	0.55	0.09	0.74	0.76	0.09	0.27	0.59
C18	.	.	.	.	.	.	.	.	.	.	.	.
C19	0.57	0.66	0.58	0.51	0.79	0.21	-1.26	.	0.72	0.17	.	0.58
C20	1.43	0.71	0.38	0.53	0.66	0.47	-0.44	0.51	0.69	0.12	0.85	0.49
C22	0.83	0.84	0.75	0.71	0.77	0.72	-1.20	.	0.81	0.58	.	0.78
C23	1.90	0.77	0.53	0.61	0.71	0.63	0.00	0.59	0.73	0.28	0.67	0.40
C24	2.00	0.74	0.48	0.51	0.62	0.64	-0.09	0.55	0.71	0.13	0.61	0.33
C26	.	.	.	.	.	.	.	.	.	.	.	.
C33	.	.	.	.	.	.	.	.	.	.	.	.
C34	.	.	.	.	.	.	.	.	.	.	.	.
U01	0.90	-0.83	-1.96	-2.17	-0.99	-0.72	-1.98	0.12	0.13	-2.28	-0.90	-0.99
U02	.	.	.	.	.	.	.	.	.	.	.	.
U03	1.20	-0.21	-0.28	-0.67	-0.11	0.00	-0.46	0.52	0.55	-0.70	0.14	0.41
U04	0.57	-2.65	-3.29	-3.91	-1.88	-0.96	-2.19	-0.27	-0.02	-2.04	-1.91	-2.62
U05	1.10	0.08	-1.44	-0.46	0.04	-0.17	-1.37	0.17	0.26	-1.37	-0.12	-0.20
U06	0.67	-1.79	-2.52	-3.32	-1.34	-0.50	-1.53	-0.09	0.11	-1.77	-0.60	-2.27
U07	.	.	.	.	.	.	.	.	.	.	.	.
U08	0.67	-1.07	-0.35	-1.49	-0.63	-0.56	-1.59	0.25	0.38	-1.31	-0.44	0.28
U09	.	.	.	.	.	.	.	.	.	.	.	.
U10	1.37	0.31	-0.74	-0.16	0.14	0.20	-1.02	0.27	0.38	-0.80	0.33	-0.24
U11	0.93	0.28	-1.04	-0.34	-0.08	-0.04	-1.72	-0.04	0.19	-1.09	-0.04	-0.67
U12	.	.	.	.	.	.	.	.	.	.	.	.
U13	1.57	0.20	-0.02	-0.06	0.39	0.24	-0.17	0.65	0.69	-0.23	0.25	0.62
U14	1.17	0.28	0.22	0.24	0.60	0.10	-0.40	0.59	0.66	-0.30	0.51	0.65
U15	.	.	.	.	.	.	.	.	.	.	.	.
U16	.	.	.	.	.	.	.	.	.	.	.	.
U17	0.77	-1.24	-2.30	-2.39	-0.88	-1.20	-2.52	-0.08	-0.06	-3.07	-1.77	-1.03
U18	.	.	.	.	.	.	.	.	.	.	.	.
U19	0.93	-0.34	-1.52	-1.02	-0.63	0.53	-0.83	0.01	0.55	-0.61	0.11	-1.14
U20	.	.	.	.	.	.	.	.	.	.	.	.

TABLE D.10 CONTINUED

STANDARDIZED RESIDUAL = (VALIDATION-MODEL)/VALIDATION VALUE

FALL

		MODELS										
ISDME CODE	VALIDATION SITE	TVA	AES	SIMPMOD	STATMOD	MOE	SERTAD	RTM-II	RIVAD	ASTRAP	UMACID	RELMAP
U21	1.43	0.01	-0.35	-0.44	0.10	0.63	0.02	0.54	0.67	0.01	0.18	-0.06
U22	1.52	0.19	-0.13	-0.21	0.24	0.75	0.08	0.53	0.72	0.15	0.42	-0.00
U23	1.30	-0.41	-0.21	-0.88	-0.25	-0.20	-0.55	0.51	0.45	-0.88	-0.25	0.17
U24	1.13	-1.07	-1.29	-2.39	-0.99	-0.67	-0.99	0.28	0.29	-2.40	-0.96	-0.62
U25	0.73	0.40	0.39	0.36	0.71	-0.19	-0.17	0.58	0.67	-0.06	0.54	0.86
U26	0.90	-0.82	-1.70	-2.38	-1.42	0.17	-1.22	0.14	0.36	-1.31	0.26	-1.33
U40	.	.	.	.	.	.	.	.	.	.	.	.
U41	1.27	-0.07	-0.88	-0.86	-0.26	0.03	-1.00	0.33	0.38	-1.25	0.11	-0.18
U42	1.42	-0.09	-0.91	-0.74	-0.07	-0.09	-0.78	0.42	0.42	-1.15	-0.33	-0.14
U43	0.83	-0.76	-2.30	-2.65	-2.01	-0.19	-1.56	-0.05	0.21	-2.10	-0.28	-1.63
U44	2.03	-0.40	-0.39	-1.21	-0.41	0.18	0.06	0.64	0.62	-0.26	-0.36	-0.39
U45	1.23	-0.71	-0.54	-1.14	-0.29	-0.62	-0.68	0.42	0.41	-1.34	-0.24	-0.09
U46	1.37	-1.06	-0.47	-1.39	-0.71	-0.47	-0.62	0.44	0.47	-1.69	-0.38	-0.71
U47	0.93	-1.08	-0.65	-1.75	-0.78	-0.26	-0.88	0.34	0.37	-1.82	-0.35	-0.12
U48	1.07	-0.80	-0.39	-1.47	-1.01	0.08	-0.63	0.38	0.53	-1.02	-0.28	-0.04
U49	.	.	.	.	.	.	.	.	.	.	.	.
U50	.	.	.	.	.	.	.	.	.	.	.	.
U51	.	.	.	.	.	.	.	.	.	.	.	.
U52	.	.	.	.	.	.	.	.	.	.	.	.
U53	.	.	.	.	.	.	.	.	.	.	.	.
U54	.	.	.	.	.	.	.	.	.	.	.	.
U55	1.80	0.14	-0.07	-0.32	0.02	0.59	0.09	0.64	0.70	0.20	0.48	0.15
U56	.	.	.	.	.	.	.	.	.	.	.	.
U57	.	.	.	.	.	.	.	.	.	.	.	.
U58	.	.	.	.	.	.	.	.	.	.	.	.
U59	.	.	.	.	.	.	.	.	.	.	.	.

TABLE D.11 STATISTICAL COMPARISON OF THE OBSERVATIONS AND PREDICTIONS OF ANNUAL 1980 SULFUR WET DEPOSITION.

TOTAL WET SULFUR DEPOSITION (kg/ha)

COMPARISON OF ALL THE MODELS PREDICTED VALUES
AGAINST THE OBSERVED VALIDATION SITES VALUES

ANNUAL 1980

				MEAN						STANDARD DEVIATION						CORRELATION OBSERVED & PREDICTED		
	NUMBER OF SITES			ALL		HIGH		LOW		ALL		HIGH		LOW				
MODEL	ALL	HIGH	LOW	OBS	PRED	OBS	PRED	OBS	PRED	OBS	PRED	OBS	PRED	OBS	PRED	ALL	HIGH	LOW
AES	32	15	17	8.66	9.08	11.43	10.74	6.23	7.61	3.48	4.34	2.46	3.07	2.16	4.83	0.51	0.39	0.41
ASTRAP	32	15	17	8.66	11.20	11.43	14.78	6.23	8.04	3.48	5.30	2.46	4.04	2.16	4.19	0.67	0.27	0.47
MOE	32	15	17	8.66	6.50	11.43	9.39	6.23	3.95	3.48	3.37	2.46	1.97	2.16	1.97	0.75	0.42	0.29
RELMAP	32	15	17	8.66	11.53	11.43	14.59	6.23	8.83	3.48	7.65	2.46	8.19	2.16	6.16	0.51	0.36	0.35
RIVAD	32	15	17	8.66	4.31	11.43	5.40	6.23	3.34	3.48	1.57	2.46	0.86	2.16	1.41	0.71	0.36	0.49
RTM	30	15	15	9.03	4.82	11.43	5.24	6.64	4.40	3.27	1.08	2.46	0.97	1.95	1.05	0.50	0.32	0.36
SERTAD	32	15	17	8.66	9.24	11.43	10.80	6.23	7.87	3.48	2.25	2.46	1.32	2.16	2.01	0.72	0.44	0.50
SIMPMOD	32	15	17	8.66	8.95	11.43	11.21	6.23	6.95	3.48	4.60	2.46	3.69	2.16	4.48	0.55	0.32	0.35
STATMOD	32	15	17	8.66	6.05	11.43	7.99	6.23	4.34	3.48	3.59	2.46	3.35	2.16	2.91	0.61	0.40	0.38
TVA	32	15	17	8.66	8.93	11.43	12.03	6.23	6.19	3.48	4.83	2.46	3.82	2.16	3.92	0.63	0.31	0.35
UMACID	30	15	15	9.03	6.40	11.43	8.42	6.64	4.39	3.27	3.20	2.46	2.19	1.95	2.78	0.66	0.63	0.10

	RESIDUAL = VALIDATION - MODEL						STANDARDIZED RESIDUAL = RESIDUAL/VALIDATION						CORRELATION RESIDUAL & PREDICTED		
	MEAN			STD DEV			MEAN			STD DEV					
MODEL	ALL	HIGH	LOW	ALL	HIGH	LOW	ALL	HIGH	LOW	ALL	HIGH	LOW	ALL	HIGH	LOW
AES	-0.42	0.68	-1.38	3.94	3.11	4.41	-0.08	0.04	-0.18	0.59	0.26	0.76	-0.65	-0.68	-0.89
ASTRAP	-2.53	-3.36	-1.81	3.92	4.12	3.71	-0.30	-0.33	-0.28	0.54	0.42	0.64	-0.76	-0.82	-0.86
MOE	2.16	2.03	2.28	2.41	2.42	2.46	0.25	0.16	0.33	0.29	0.19	0.34	-0.31	-0.39	-0.54
RELMAP	-2.86	-3.16	-2.60	6.61	7.65	5.77	-0.33	-0.28	-0.36	0.81	0.58	0.98	-0.89	-0.95	-0.94
RIVAD	4.36	6.03	2.88	2.61	2.29	1.92	0.48	0.51	0.45	0.18	0.11	0.22	0.35	0.01	-0.18
RTM	4.21	6.18	2.24	2.88	2.33	1.85	0.41	0.53	0.30	0.19	0.11	0.19	0.20	-0.07	-0.19
SERTAD	-0.58	0.63	-1.64	2.41	2.22	2.10	-0.18	0.03	-0.36	0.36	0.18	0.39	0.11	-0.11	-0.45
SIMPMOD	-0.28	0.21	-0.72	3.97	3.73	4.24	-0.06	-0.01	-0.10	0.56	0.34	0.70	-0.68	-0.78	-0.88
STATMOD	2.62	3.44	1.89	3.13	3.28	2.89	0.31	0.29	0.32	0.37	0.27	0.45	-0.47	-0.72	-0.72
TVA	-0.26	-0.60	0.04	3.76	3.86	3.76	-0.03	-0.08	0.01	0.51	0.34	0.62	-0.70	[illegible]	-0.84
UMACID	2.63	3.01	2.25	2.67	2.01	3.23	0.29	0.26	0.32	0.37	0.16	0.50	-0.39	-0.32	-0.80

TABLE D.12 STATISTICAL COMPARISON OF THE OBSERVATIONS AND PREDICTIONS OF WINTER 1980 SULFUR WET DEPOSITION.

TOTAL WET SULFUR DEPOSITION (kg/ha)

COMPARISON OF ALL THE MODELS PREDICTED VALUES
AGAINST THE OBSERVED VALIDATION SITES VALUES

WINTER 1980

MODEL	NUMBER OF SITES			MEAN						STANDARD DEVIATION						CORRELATION OBSERVED & PREDICTED		
				ALL		HIGH		LOW		ALL		HIGH		LOW				
	ALL	HIGH	LOW	OBS	PRED	OBS	PRED	OBS	PRED	OBS	PRED	OBS	PRED	OBS	PRED	ALL	HIGH	LOW
AES	38	15	23	1.40	2.26	1.73	2.50	1.19	2.10	0.70	1.53	0.63	1.23	0.66	1.70	0.40	-0.13	0.62
ASTRAP	38	15	23	1.40	2.11	1.73	2.80	1.19	1.67	0.70	0.97	0.63	0.65	0.66	0.88	0.50	-0.01	0.55
MOE	38	15	23	1.40	1.07	1.73	1.56	1.19	0.75	0.70	0.56	0.63	0.34	0.66	0.44	0.38	0.02	0.24
RELMAP	38	15	23	1.40	1.50	1.73	1.82	1.19	1.29	0.70	0.93	0.63	0.77	0.66	0.98	0.40	-0.06	0.52
RIVAD	38	15	23	1.40	0.77	1.73	0.95	1.19	0.65	0.70	0.39	0.63	0.31	0.66	0.40	0.43	-0.18	0.58
RTM	36	15	21	1.46	0.99	1.73	1.02	1.26	0.97	0.68	0.42	0.63	0.34	0.65	0.48	0.39	-0.20	0.70
SERTAD	38	15	23	1.40	1.40	1.73	1.86	1.19	1.10	0.70	0.75	0.63	0.66	0.66	0.66	0.44	0.21	0.37
SIMPMOD	38	15	23	1.40	2.80	1.73	3.60	1.19	2.27	0.70	1.54	0.63	1.27	0.66	1.50	0.46	-0.00	0.53
STATMOD	38	15	23	1.40	1.66	1.73	1.95	1.19	1.46	0.70	1.08	0.63	0.94	0.66	1.14	0.54	-0.08	0.80
TVA	38	15	23	1.40	1.49	1.73	2.03	1.19	1.13	0.70	0.83	0.63	0.66	0.66	0.74	0.48	-0.03	0.56
UMACID	35	15	20	1.48	1.45	1.73	1.75	1.29	1.23	0.67	0.77	0.63	0.45	0.64	0.88	0.45	0.25	0.46

MODEL	RESIDUAL = VALIDATION - MODEL						STANDARDIZED RESIDUAL = RESIDUAL/VALIDATION						CORRELATION RESIDUAL & PREDICTED		
	MEAN			STD DEV			MEAN			STD DEV					
	ALL	HIGH	LOW	ALL	HIGH	LOW	ALL	HIGH	LOW	ALL	HIGH	LOW	ALL	HIGH	LOW
AES	-0.85	-0.77	-0.91	1.40	1.45	1.40	-0.67	-0.62	-0.70	1.25	0.85	1.46	-0.89	-0.90	-0.93
ASTRAP	-0.71	-1.06	-0.48	0.86	0.92	0.76	-0.65	-0.80	-0.55	0.81	0.70	0.88	-0.72	-0.72	-0.69
MOE	0.34	0.18	0.44	0.71	0.72	0.70	0.10	0.00	0.16	0.52	0.35	0.60	-0.42	-0.46	-0.40
RELMAP	-0.09	-0.08	-0.10	0.91	1.03	0.85	-0.11	-0.16	-0.07	0.76	0.58	0.87	-0.71	-0.79	-0.75
RIVAD	0.64	0.78	0.54	0.63	0.76	0.54	0.40	0.38	0.41	0.32	0.26	0.36	-0.15	-0.57	-0.03
RTM	0.47	0.72	0.29	0.64	0.78	0.46	0.24	0.34	0.16	0.35	0.29	0.38	-0.24	-0.60	-0.06
SERTAD	0.01	-0.12	0.09	0.77	0.81	0.74	-0.10	-0.16	-0.05	0.65	0.53	0.72	-0.58	-0.65	-0.57
SIMPMOD	-1.39	-1.87	-1.08	1.37	1.42	1.28	-1.11	-1.30	-0.98	1.21	1.00	1.34	-0.89	-0.89	-0.90
STATMOD	-0.25	-0.22	-0.27	0.92	1.18	0.73	-0.17	-0.25	-0.12	0.74	0.67	0.79	-0.77	-0.84	-0.84
TVA	-0.08	-0.29	0.06	0.79	0.93	0.66	-0.12	-0.29	-0.01	0.63	0.54	0.67	-0.63	-0.73	-0.56
UMACID	0.03	-0.02	0.06	0.75	0.68	0.82	-0.04	-0.10	-0.00	0.70	0.36	0.88	-0.61	-0.43	-0.72

TABLE D.13 STATISTICAL COMPARISON OF THE OBSERVATIONS AND PREDICTIONS OF SPRING 1980 SULFUR WET DEPOSITION.

TOTAL WET SULFUR DEPOSITION (kg/ha)

COMPARISON OF ALL THE MODELS PREDICTED VALUES
AGAINST THE OBSERVED VALIDATION SITES VALUES

SPRING 1980

	NUMBER OF SITES			MEAN						STANDARD DEVIATION						CORRELATION OBSERVED & PREDICTED		
				ALL		HIGH		LOW		ALL		HIGH		LOW				
MODEL	ALL	HIGH	LOW	OBS	PRED	OBS	PRED	OBS	PRED	OBS	PRED	OBS	PRED	OBS	PRED	ALL	HIGH	LOW
AES	46	15	31	2.51	1.93	3.45	2.69	2.06	1.56	1.04	1.06	0.67	0.70	0.87	1.02	0.71	0.48	0.61
ASTRAP	46	15	31	2.51	2.69	3.45	3.82	2.06	2.15	1.04	1.42	0.67	0.97	0.87	1.29	0.66	0.26	0.53
MOE	46	15	31	2.51	1.87	3.45	2.90	2.06	1.38	1.04	1.06	0.67	0.71	0.87	0.81	0.73	0.39	0.57
RELMAP	43	15	28	2.60	3.42	3.45	4.83	2.15	2.66	1.00	2.39	0.67	2.85	0.85	1.72	0.66	0.72	0.49
RIVAD	46	15	31	2.51	1.24	3.45	1.68	2.06	1.03	1.04	0.51	0.67	0.20	0.87	0.47	0.72	0.18	0.61
RTM	42	15	27	2.65	1.29	3.45	1.52	2.21	1.16	0.96	0.35	0.67	0.27	0.81	0.33	0.59	0.36	0.44
SERTAD	46	15	31	2.51	2.16	3.45	2.81	2.06	1.85	1.04	0.81	0.67	0.65	0.87	0.68	0.74	0.44	0.66
SIMPMOD	45	15	30	2.53	1.99	3.45	2.73	2.07	1.62	1.04	1.06	0.67	0.82	0.88	0.97	0.73	0.63	0.62
STATMOD	45	15	30	2.53	1.48	3.45	2.12	2.07	1.16	1.04	0.93	0.67	0.88	0.88	0.79	0.76	0.53	0.74
TVA	46	15	31	2.51	2.06	3.45	3.03	2.06	1.60	1.04	1.18	0.67	0.91	0.87	1.00	0.76	0.59	0.64
UMACID	41	15	26	2.68	1.38	3.45	2.01	2.23	1.02	0.96	0.82	0.67	0.61	0.82	0.70	0.72	0.65	0.53

	RESIDUAL = VALIDATION - MODEL						STANDARDIZED RESIDUAL = RESIDUAL/VALIDATION						CORRELATION RESIDUAL & PREDICTED		
	MEAN			STD DEV			MEAN			STD DEV					
MODEL	ALL	HIGH	LOW	ALL	HIGH	LOW	ALL	HIGH	LOW	ALL	HIGH	LOW	ALL	HIGH	LOW
AES	0.58	0.76	0.49	0.80	0.70	0.84	0.24	0.21	0.25	0.46	0.19	0.55	-0.40	-0.54	-0.58
ASTRAP	-0.18	-0.37	-0.09	1.08	1.02	1.11	-0.08	-0.13	-0.06	0.57	0.31	0.66	-0.69	-0.77	-0.75
MOE	0.64	0.55	0.68	0.77	0.77	0.79	0.27	0.15	0.32	0.37	0.22	0.41	-0.39	-0.59	-0.41
RELMAP	-0.82	-1.39	-0.51	1.89	2.41	1.51	-0.29	-0.36	-0.26	0.75	0.55	0.85	-0.92	-0.98	-0.87
RIVAD	1.27	1.77	1.03	0.76	0.67	0.69	0.49	0.50	0.48	0.22	0.10	0.26	0.31	-0.12	0.08
RTM	1.36	1.93	1.05	0.81	0.63	0.73	0.47	0.55	0.43	0.21	0.09	0.25	0.27	-0.04	0.03
SERTAD	0.35	0.64	0.21	0.70	0.70	0.66	0.07	0.17	0.02	0.36	0.20	0.41	-0.05	-0.51	-0.17
SIMPMOD	0.54	0.72	0.45	0.77	0.65	0.82	0.21	0.21	0.21	0.43	0.19	0.52	-0.38	-0.60	-0.53
STATMOD	1.05	1.33	0.91	0.69	0.77	0.61	0.44	0.39	0.47	0.30	0.23	0.33	-0.19	-0.67	-0.22
TVA	0.45	0.42	0.46	0.78	0.75	0.80	0.19	0.12	0.22	0.45	0.23	0.53	-0.50	-0.69	-0.55
UMACID	1.30	1.44	1.22	0.68	0.54	0.74	0.50	0.42	0.54	0.29	0.13	0.35	-0.18	-0.31	-0.36

TABLE D.14 STATISTICAL COMPARISON OF THE OBSERVATIONS AND PREDICTIONS OF SUMMER 1980 SULFUR WET DEPOSITION.

TOTAL WET SULFUR DEPOSITION (kg/ha)

COMPARISON OF ALL THE MODELS PREDICTED VALUES
AGAINST THE OBSERVED VALIDATION SITES VALUES

SUMMER 1980

MODEL	NUMBER OF SITES			MEAN						STANDARD DEVIATION						CORRELATION OBSERVED & PREDICTED		
				ALL		HIGH		LOW		ALL		HIGH		LOW				
	ALL	HIGH	LOW	OBS	PRED	OBS	PRED	OBS	PRED	OBS	PRED	OBS	PRED	OBS	PRED	ALL	HIGH	LOW
AES	45	15	30	2.99	2.62	4.60	3.69	2.19	2.09	1.72	1.45	1.43	0.97	1.23	1.36	0.67	0.46	0.53
ASTRAP	45	15	30	2.99	3.44	4.60	5.43	2.19	2.44	1.72	2.19	1.43	1.97	1.23	1.53	0.72	0.54	0.48
MOE	45	15	30	2.99	2.13	4.60	3.36	2.19	1.52	1.72	1.21	1.43	0.76	1.23	0.88	0.79	0.55	0.63
RELMAP	44	15	29	3.05	4.45	4.60	6.61	2.26	3.33	1.69	3.56	1.43	4.10	1.20	2.69	0.54	0.41	0.32
RIVAD	45	15	30	2.99	1.55	4.60	2.09	2.19	1.28	1.72	0.63	1.43	0.35	1.23	0.55	0.76	0.66	0.59
RTM	42	15	27	3.16	1.64	4.60	1.99	2.35	1.44	1.66	0.47	1.43	0.43	1.19	0.36	0.77	0.62	0.64
SERTAD	45	15	30	2.99	3.26	4.60	4.20	2.19	2.79	1.72	1.15	1.43	0.93	1.23	0.95	0.75	0.47	0.68
SIMPMOD	45	15	30	2.99	1.96	4.60	2.72	2.19	1.58	1.72	1.15	1.43	0.95	1.23	1.06	0.55	0.38	0.35
STATMOD	45	15	30	2.99	1.57	4.60	2.31	2.19	1.20	1.72	1.04	1.43	1.01	1.23	0.85	0.66	0.52	0.49
TVA	45	15	30	2.99	2.50	4.60	3.72	2.19	1.89	1.72	1.46	1.43	1.17	1.23	1.19	0.66	0.43	0.43
UMACID	41	15	26	3.21	1.94	4.60	2.93	2.41	1.37	1.65	1.23	1.43	0.91	1.17	1.00	0.63	0.51	0.31

MODEL	RESIDUAL = VALIDATION - MODEL						STANDARDIZED RESIDUAL = RESIDUAL/VALIDATION						CORRELATION RESIDUAL & PREDICTED		
	MEAN			STD DEV			MEAN			STD DEV					
	ALL	HIGH	LOW	ALL	HIGH	LOW	ALL	HIGH	LOW	ALL	HIGH	LOW	ALL	HIGH	LOW
AES	0.37	0.91	0.10	1.32	1.30	1.27	0.08	0.16	0.04	0.54	0.20	0.64	-0.23	-0.24	-0.56
ASTRAP	-0.45	-0.84	-0.25	1.53	1.69	1.42	-0.19	-0.21	-0.18	0.66	0.35	0.78	-0.62	-0.70	-0.65
MOE	0.86	1.24	0.67	1.06	1.19	0.96	0.25	0.24	0.26	0.32	0.18	0.37	0.14	0.02	-0.10
RELMAP	-1.39	-2.01	-1.07	3.01	3.75	2.57	-0.50	-0.44	-0.53	1.22	0.70	1.43	-0.88	-0.94	-0.90
RIVAD	1.44	2.50	0.91	1.31	1.23	1.01	0.41	0.52	0.35	0.29	0.09	0.33	0.52	0.48	0.17
RTM	1.52	2.60	0.92	1.34	1.21	0.99	0.39	0.55	0.30	0.22	0.10	0.22	0.60	0.38	0.40
SERTAD	-0.27	0.40	-0.60	1.14	1.29	0.91	-0.33	0.04	-0.52	0.55	0.25	0.56	0.13	-0.20	-0.12
SIMPMOD	1.03	1.88	0.61	1.45	1.38	1.31	0.28	0.38	0.23	0.51	0.21	0.60	-0.14	-0.29	-0.48
STATMOD	1.42	2.28	0.99	1.29	1.25	1.10	0.45	0.48	0.43	0.36	0.21	0.42	0.08	-0.21	-0.23
TVA	0.49	0.87	0.30	1.34	1.40	1.29	0.11	0.15	0.09	0.51	0.27	0.59	-0.24	-0.39	-0.51
UMACID	1.27	1.66	1.04	1.29	1.24	1.28	0.38	0.34	0.40	0.34	0.18	0.40	-0.14	-0.15	-0.50

TABLE D.15 STATISTICAL COMPARISON OF THE OBSERVATIONS AND PREDICTIONS OF AUTUMN 1980 SULFUR WET DEPOSITION.

TOTAL WET SULFUR DEPOSITION (kg/ha)

COMPARISON OF ALL THE MODELS PREDICTED VALUES
AGAINST THE OBSERVED VALIDATION SITES VALUES

FALL 1980

				MEAN						STANDARD DEVIATION						CORRELATION OBSERVED & PREDICTED		
	NUMBER OF SITES			ALL		HIGH		LOW		ALL		HIGH		LOW				
MODEL	ALL	HIGH	LOW	OBS	PRED	OBS	PRED	OBS	PRED	OBS	PRED	OBS	PRED	OBS	PRED	ALL	HIGH	LOW
AES	42	15	27	1.32	1.67	1.71	1.86	1.10	1.56	0.65	0.80	0.72	0.54	0.50	0.91	0.21	0.22	0.12
ASTRAP	42	15	27	1.32	2.04	1.71	2.73	1.10	1.66	0.65	0.91	0.72	0.61	0.50	0.83	0.42	0.27	0.20
MOE	42	15	27	1.32	1.11	1.71	1.58	1.10	0.85	0.65	0.50	0.72	0.32	0.50	0.37	0.35	0.22	-0.07
RELMAP	41	15	26	1.34	1.27	1.71	1.33	1.12	1.23	0.64	0.67	0.72	0.61	0.49	0.71	0.14	0.06	0.17
RIVAD	42	15	27	1.32	0.57	1.71	0.67	1.10	0.51	0.65	0.20	0.72	0.12	0.50	0.22	0.42	0.40	0.29
RTM	40	15	25	1.35	0.71	1.71	0.71	1.13	0.71	0.65	0.19	0.72	0.11	0.51	0.22	0.37	0.40	0.48
SERTAD	42	15	27	1.32	2.00	1.71	2.25	1.10	1.85	0.65	0.56	0.72	0.43	0.50	0.57	0.55	0.75	0.35
SIMPMOD	41	15	26	1.34	1.93	1.71	2.47	1.12	1.62	0.64	1.01	0.72	0.86	0.49	0.97	0.09	-0.22	-0.05
STATMOD	41	15	26	1.34	1.28	1.71	1.62	1.12	1.09	0.64	0.67	0.72	0.55	0.49	0.67	0.11	-0.26	0.05
TVA	42	15	27	1.32	1.28	1.71	1.77	1.10	1.01	0.65	0.70	0.72	0.57	0.50	0.62	0.15	-0.30	0.00
UMACID	38	15	23	1.39	1.27	1.71	1.72	1.18	0.98	0.64	0.60	0.72	0.51	0.49	0.46	0.40	0.54	-0.13

	RESIDUAL = VALIDATION - MODEL						STANDARDIZED RESIDUAL = RESIDUAL/VALIDATION						CORRELATION RESIDUAL & PREDICTED		
	MEAN			STD DEV			MEAN			STD DEV					
MODEL	ALL	HIGH	LOW	ALL	HIGH	LOW	ALL	HIGH	LOW	ALL	HIGH	LOW	ALL	HIGH	LOW
AES	-0.35	-0.15	-0.46	0.92	0.80	0.98	-0.45	-0.22	-0.58	0.99	0.49	1.17	-0.72	-0.48	-0.86
ASTRAP	-0.72	-1.02	-0.56	0.87	0.81	0.88	-0.72	-0.81	-0.67	0.89	0.69	1.00	-0.74	-0.52	-0.83
MOE	0.20	0.13	0.24	0.66	0.72	0.64	0.02	-0.05	0.06	0.49	0.38	0.54	-0.40	-0.22	-0.63
RELMAP	0.07	0.38	-0.10	0.86	0.92	0.79	-0.09	0.13	-0.21	0.83	0.45	0.98	-0.67	-0.63	-0.79
RIVAD	0.75	1.04	0.59	0.60	0.68	0.49	0.51	0.56	0.48	0.23	0.15	0.26	0.12	0.25	-0.15
RTM	0.64	1.00	0.42	0.61	0.68	0.45	0.38	0.52	0.30	0.26	0.16	0.28	0.09	0.27	0.04
SERTAD	-0.68	-0.54	-0.76	0.58	0.49	0.62	-0.76	-0.47	-0.93	0.69	0.47	0.75	-0.34	0.23	-0.65
SIMPMOD	-0.59	-0.76	-0.50	1.15	1.23	1.11	-0.69	-0.70	-0.67	1.18	0.83	1.35	-0.83	-0.82	-0.90
STATMOD	0.06	0.09	0.03	0.88	1.01	0.81	-0.11	-0.12	-0.10	0.76	0.56	0.87	-0.68	-0.73	-0.79
TVA	0.04	-0.06	0.09	0.88	1.04	0.80	-0.12	-0.23	-0.06	0.78	0.61	0.87	-0.68	-0.75	-0.78
UMACID	0.12	-0.01	0.21	0.68	0.61	0.72	-0.03	-0.10	0.01	0.61	0.38	0.72	-0.51	-0.20	-0.73

APPENDIX E

ANNUAL 1980 MODEL PERFORMANCES

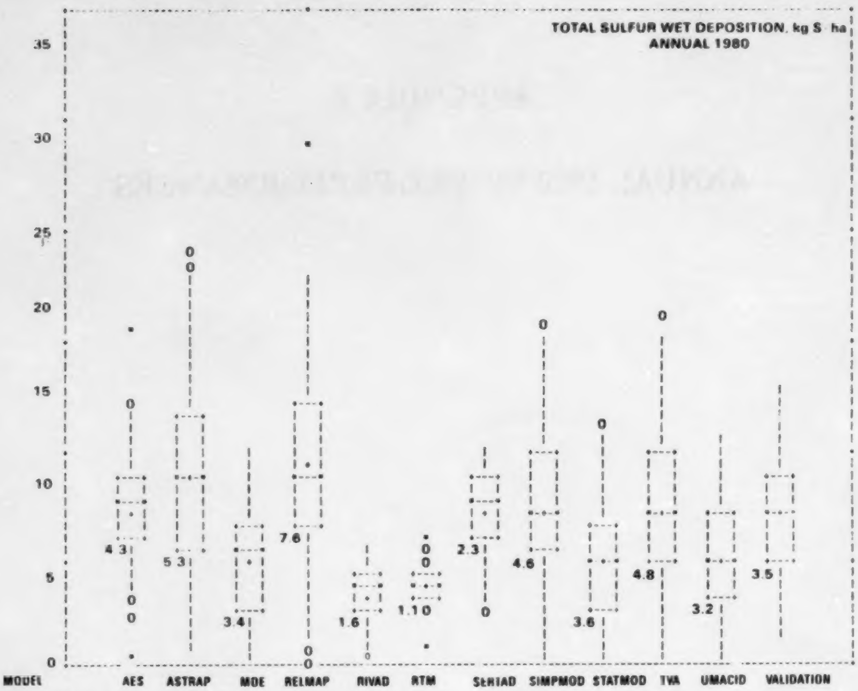


Figure E.1 **Annual 1980** box plots of the predictions and observations of sulfur wet deposition (kg S/ha).

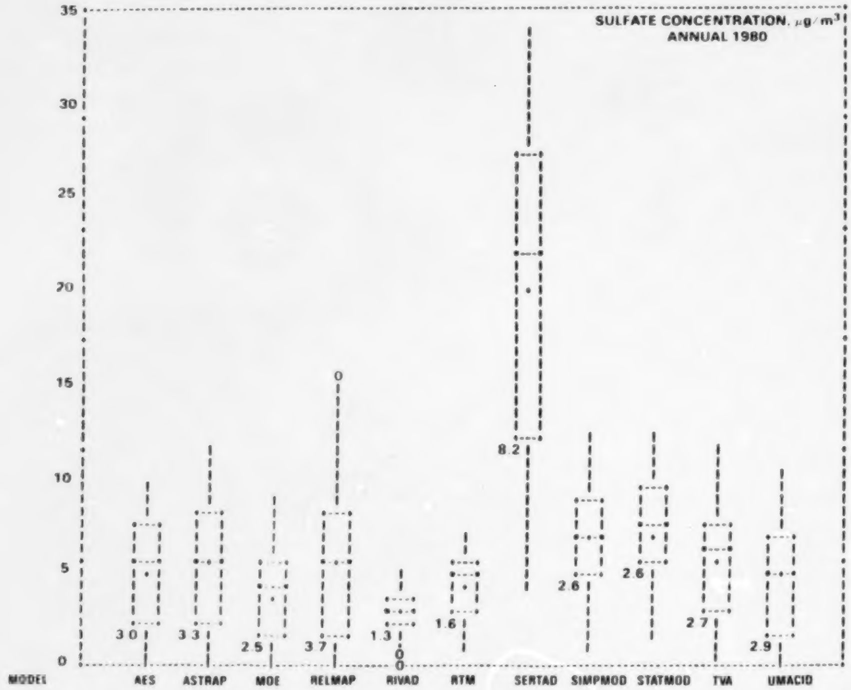


Figure E.2 **Annual 1980** box plots of the predictions of sulfate air concentrations (µg/m³).

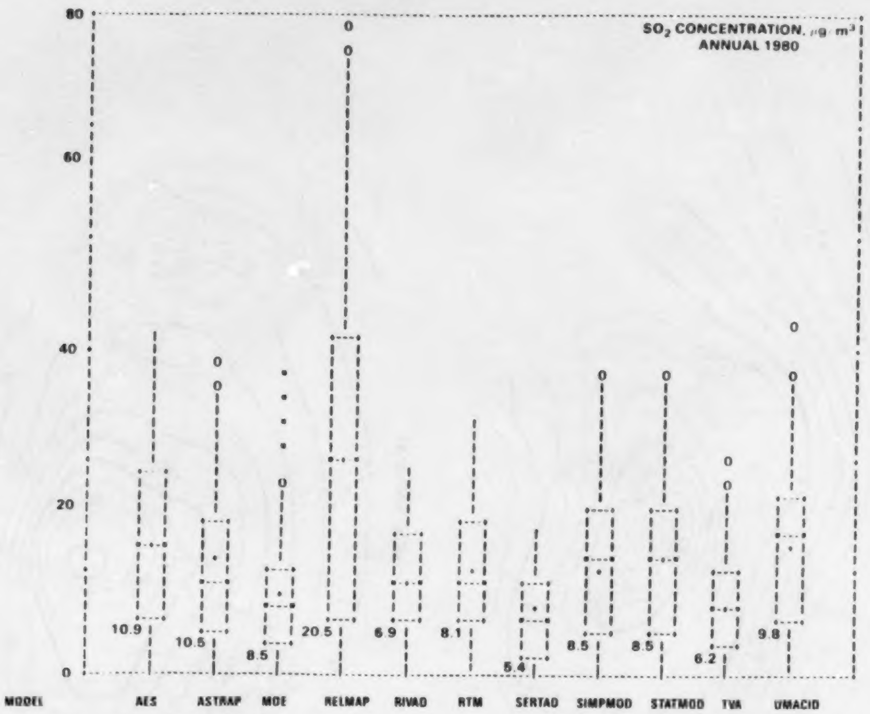


Figure E.3 **Annual 1980** box plots of the predictions of sulfur dioxide air concentrations (µg/m^3).

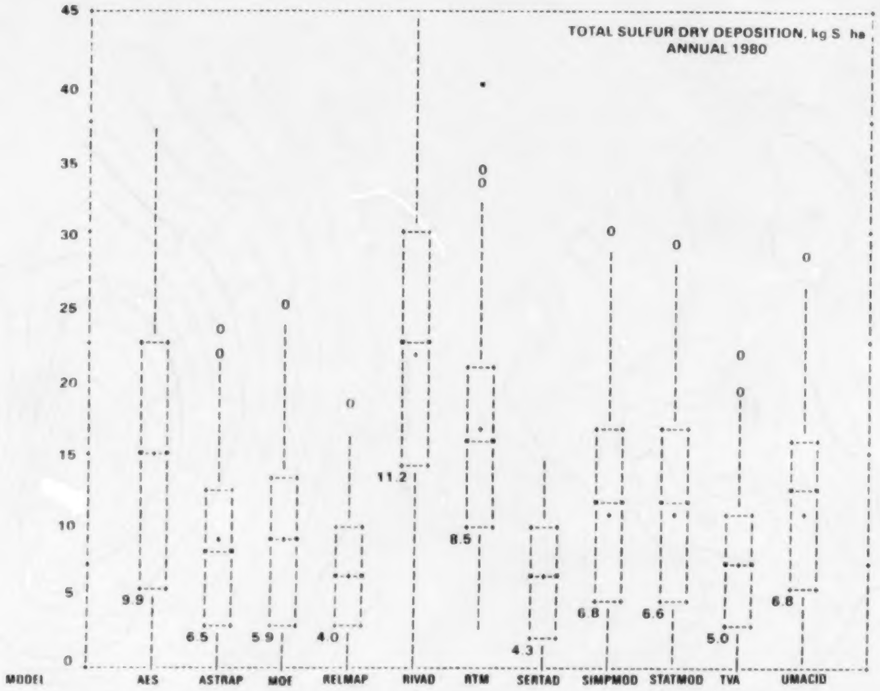


Figure E.4 **Annual 1980** box plots of the predictions of sulfur dry deposition (kg S/ha).

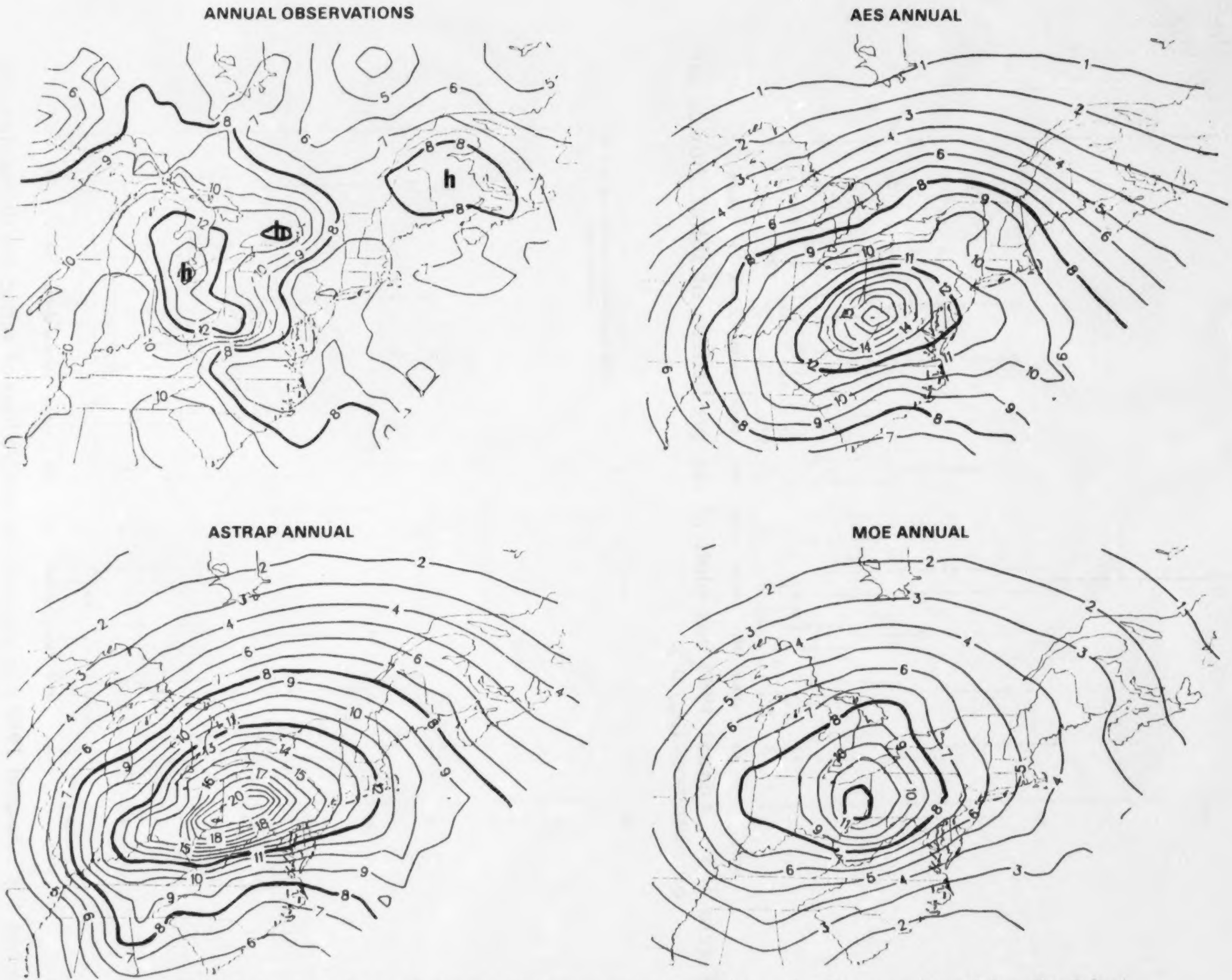
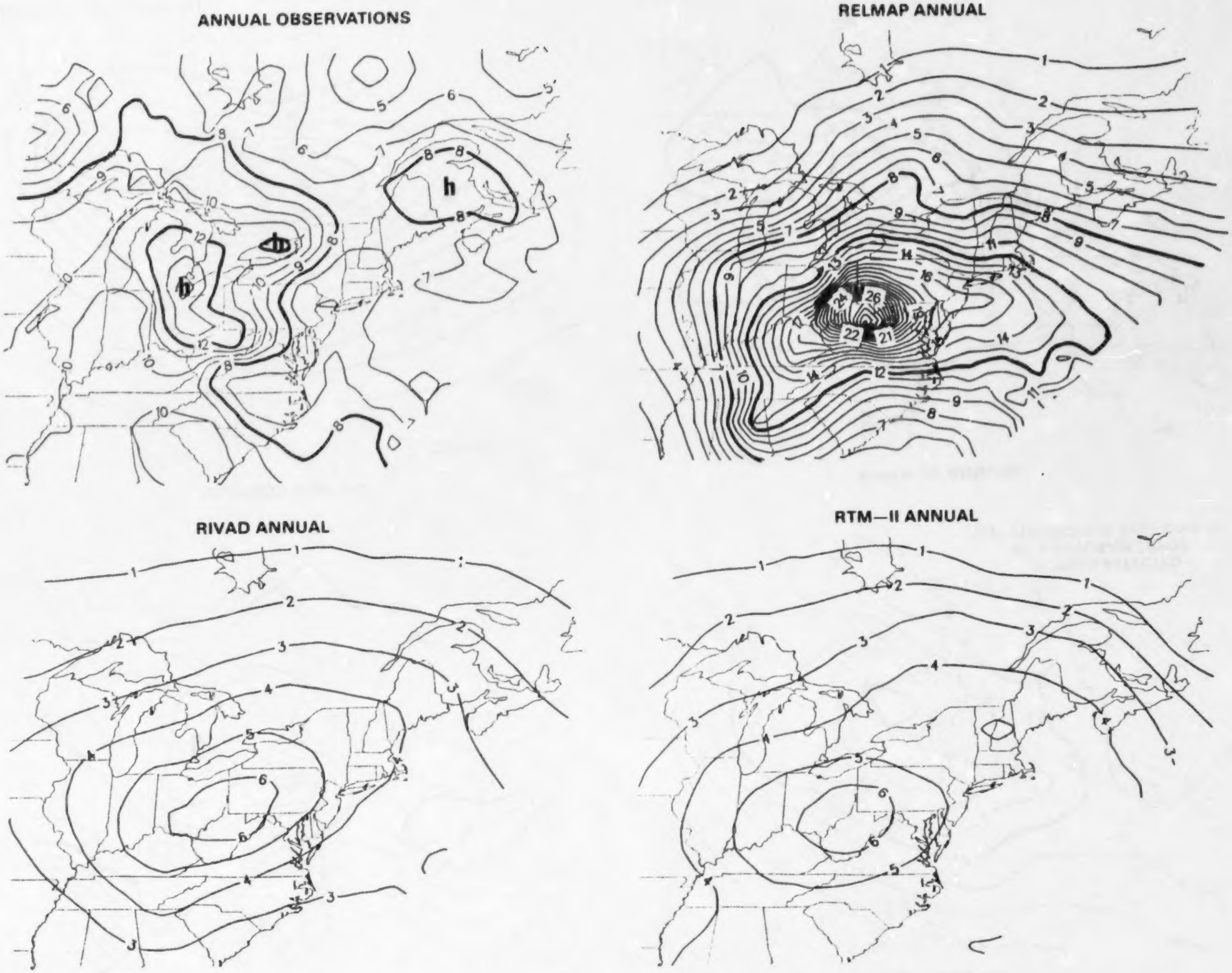


Figure E.5 **Annual 1980** kriged analyses of the predictions and observations of sulfur wet deposition (kg S/ha).



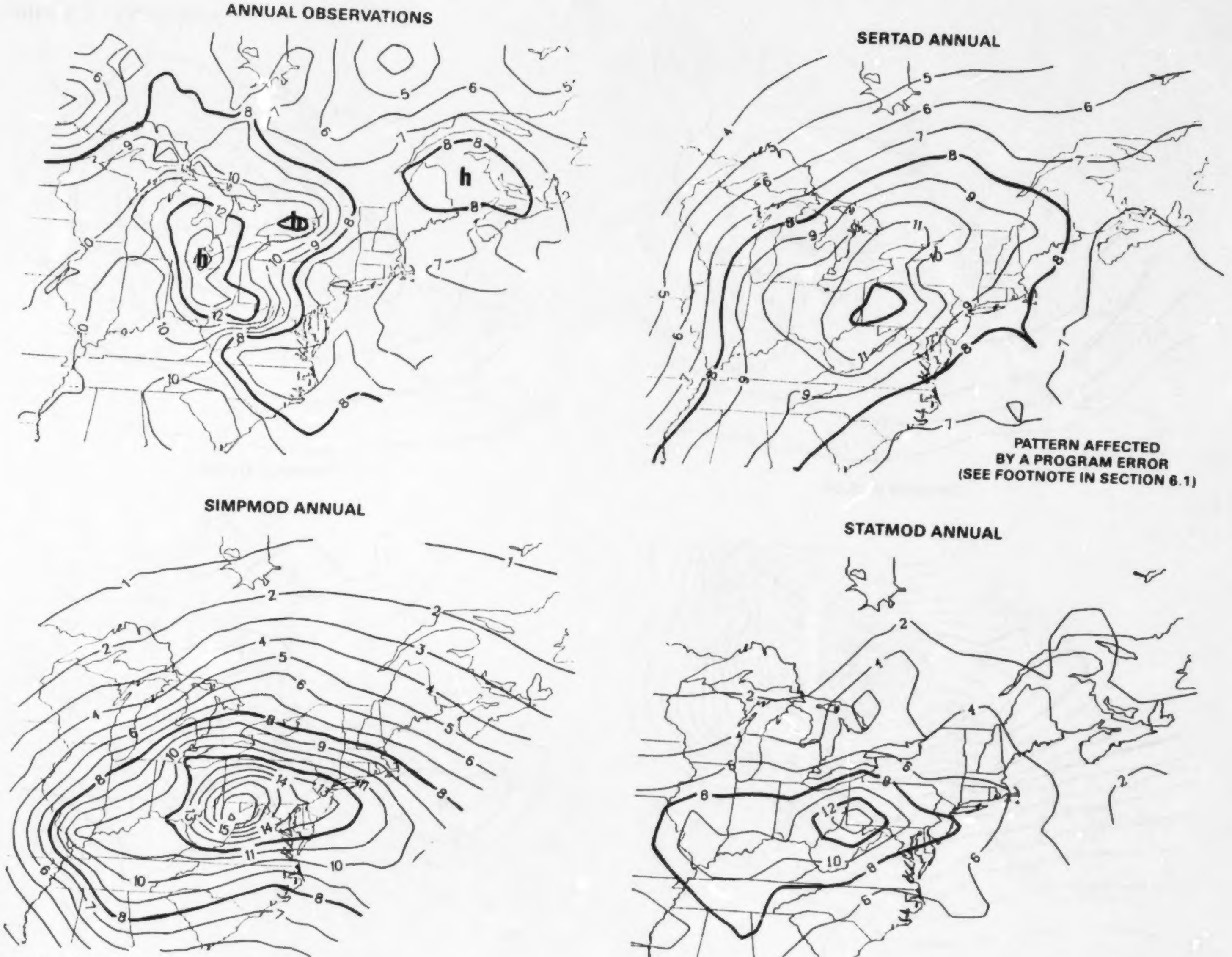
ANNUAL OBSERVATIONS

RELMAP ANNUAL

RIVAD ANNUAL

RTM—II ANNUAL

Figure E.5 Continued



ANNUAL OBSERVATIONS

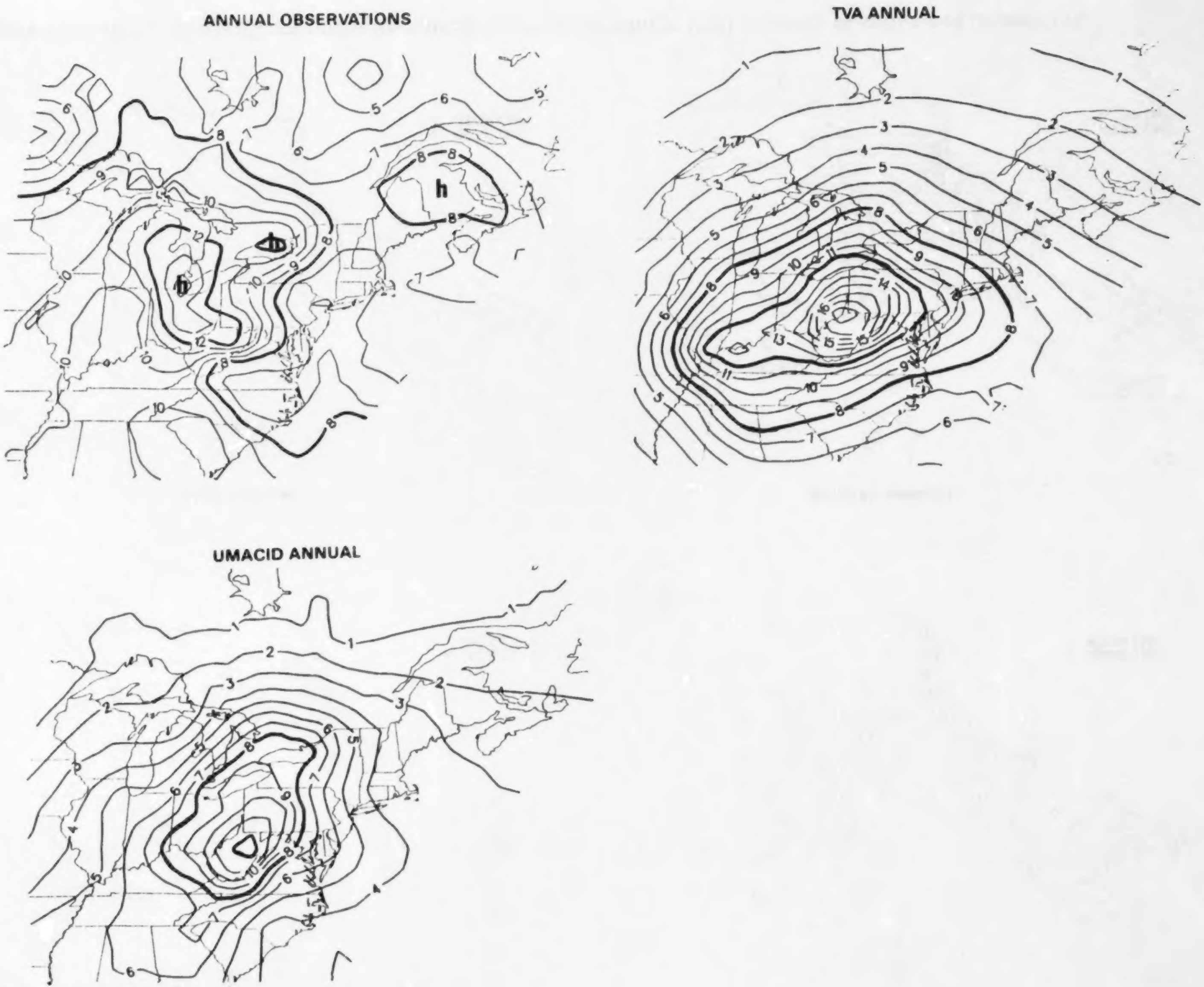
SERTAD ANNUAL

PATTERN AFFECTED
BY A PROGRAM ERROR
(SEE FOOTNOTE IN SECTION 6.1)

SIMPMOD ANNUAL

STATMOD ANNUAL

Figure E.5 Continued



ANNUAL OBSERVATIONS

TVA ANNUAL

UMACID ANNUAL

Figure E.5 Continued

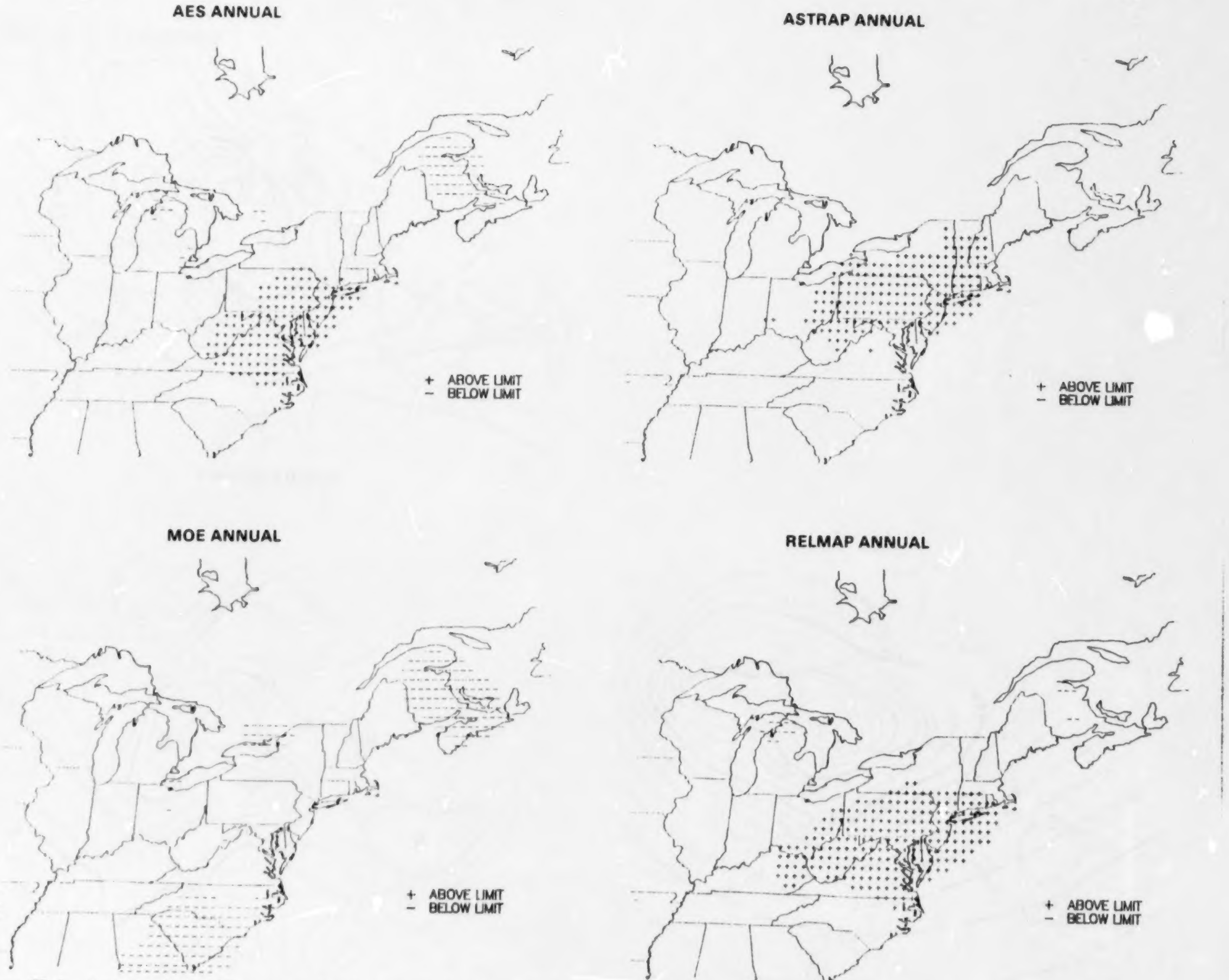
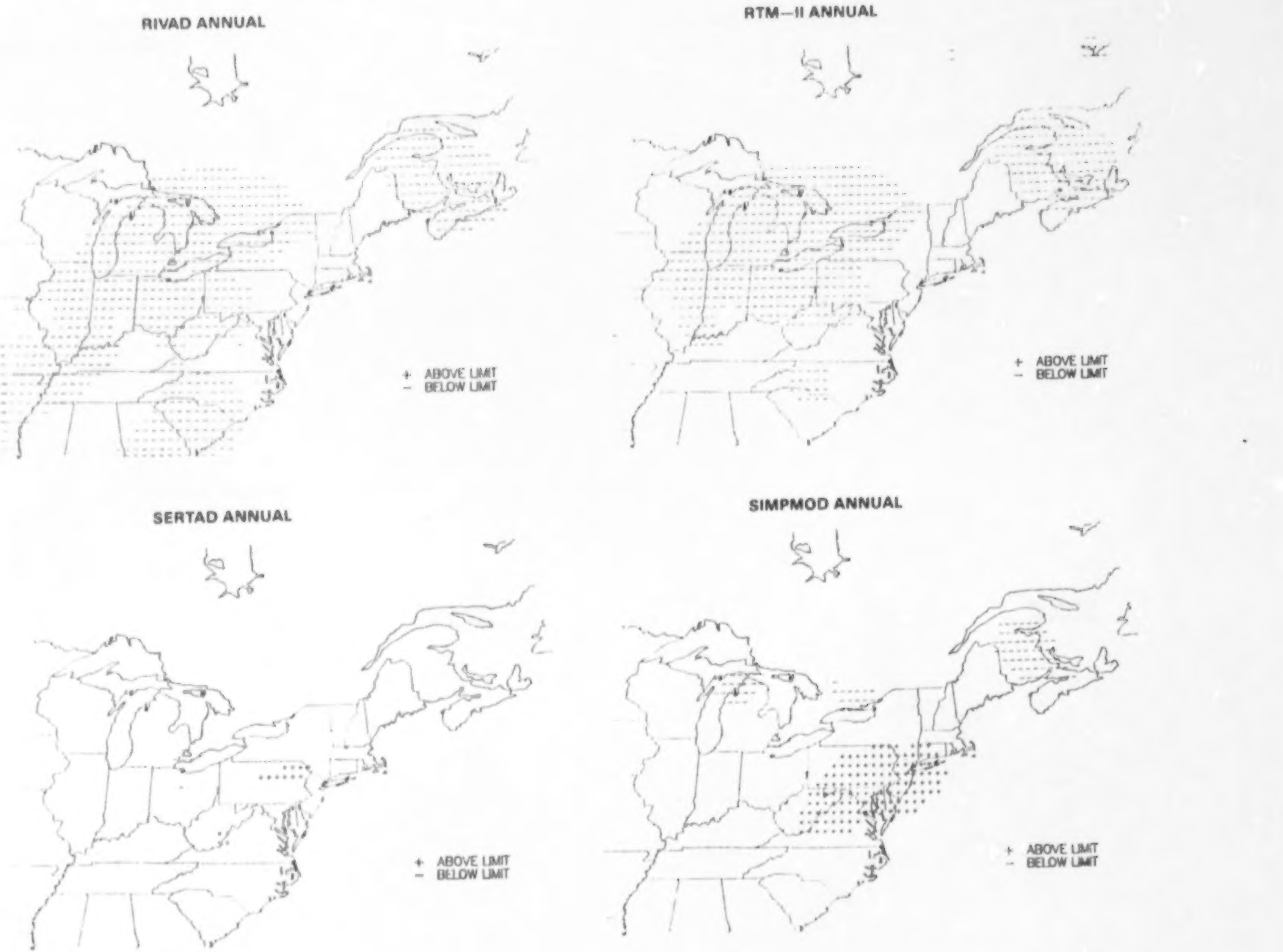


Figure E.6 Areas of significant under/overpredictions of the **winter 1980** amounts of sulfur wet deposition.


RIVAD ANNUAL

RTM—II ANNUAL

SERTAD ANNUAL

SIMPMOD ANNUAL

Figure E.6 Continued

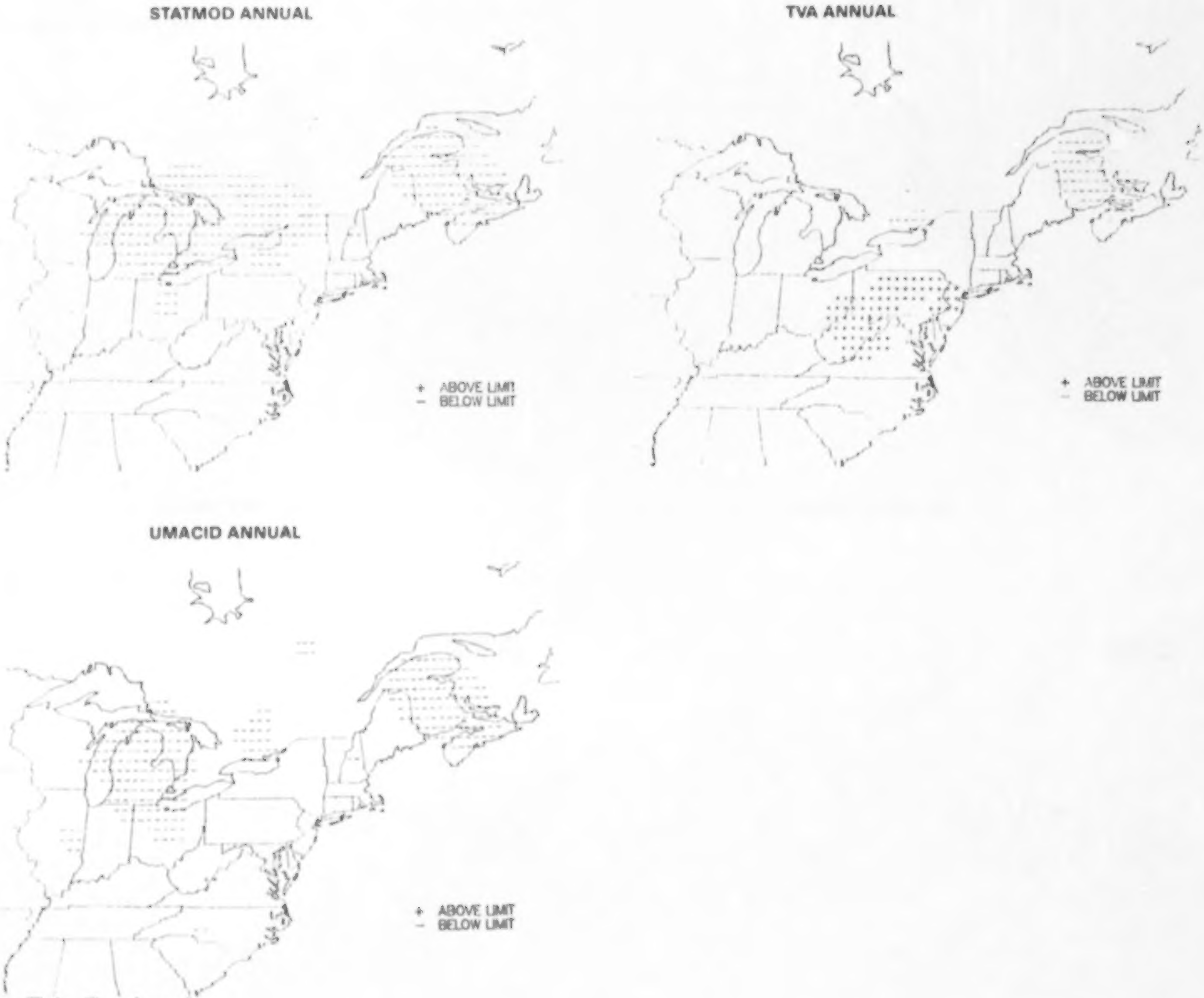


Figure E.6 Continued

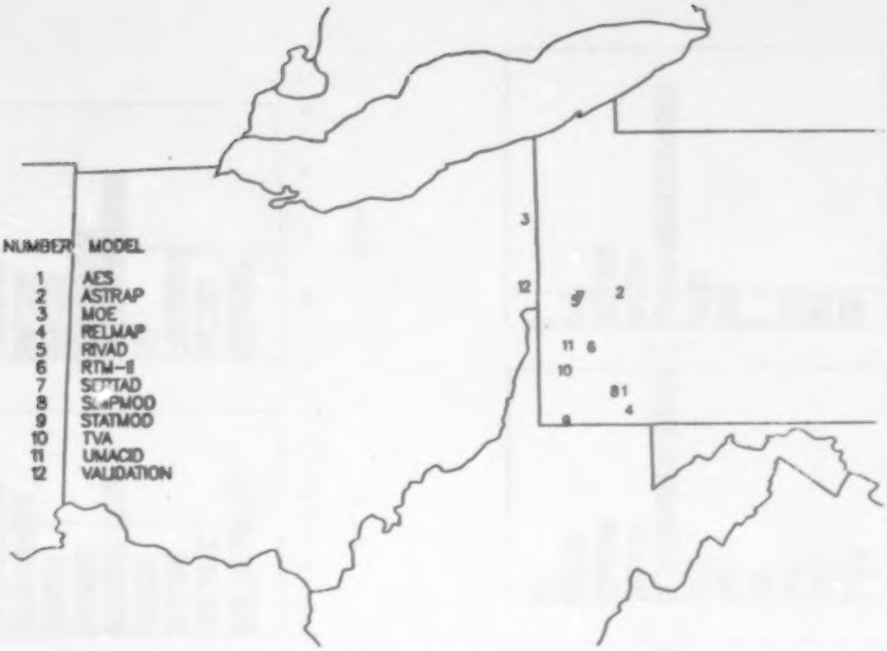


Figure E.7 **Annual 1980** locations of centroids.

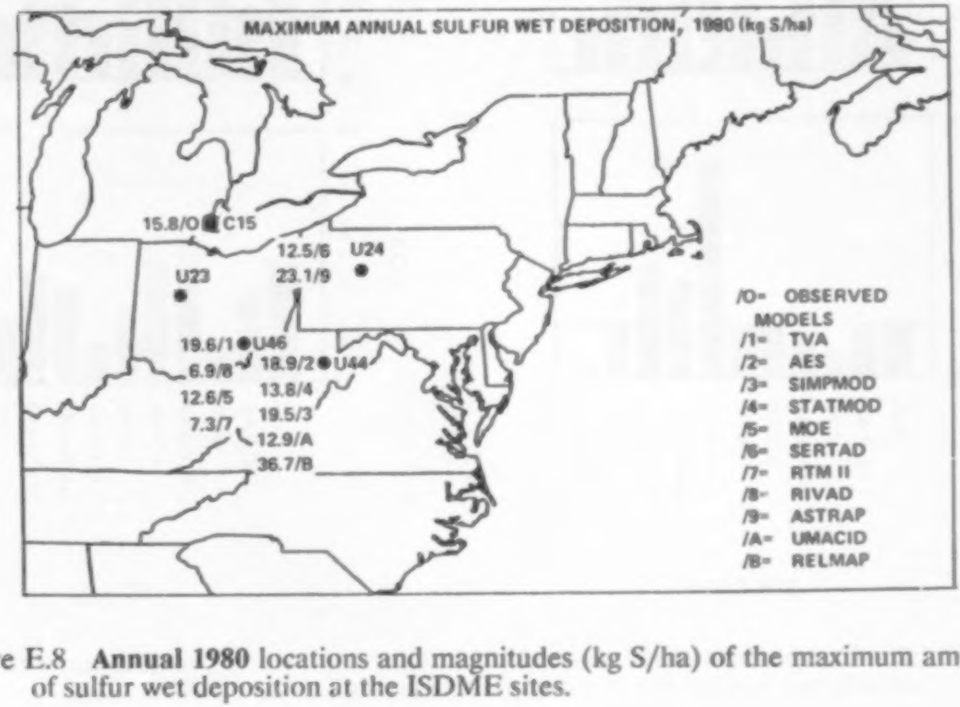


Figure E.8 **Annual 1980** locations and magnitudes (kg S/ha) of the maximum amount of sulfur wet deposition at the ISDME sites.

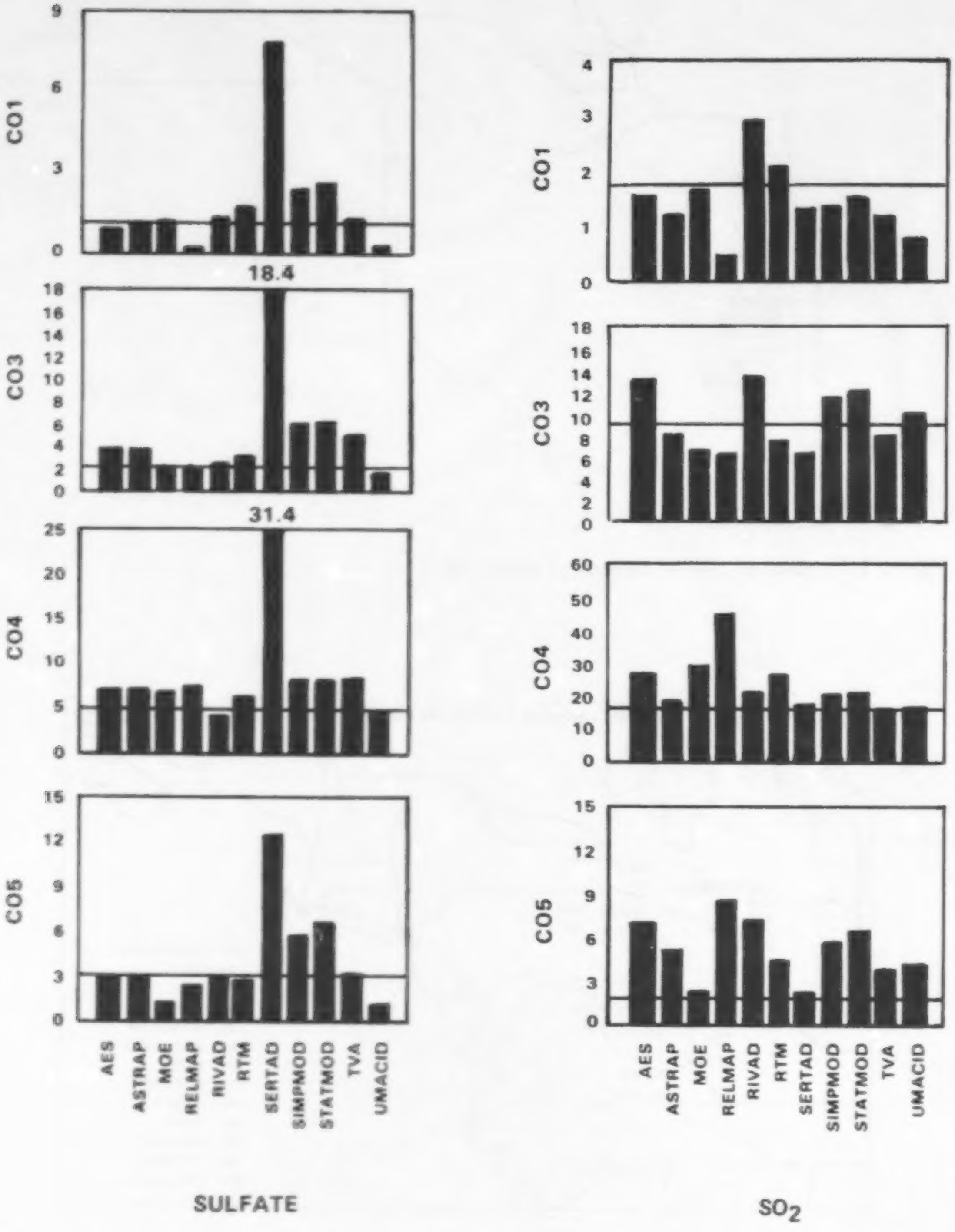


Figure E.9 Comparison of the **annual 1980** observations and predictions of the air concentrations (μg/m^3) of sulfur dioxide and sulfate at the APN sites.

TABLE E.1 SUMMARY OF MODEL PERFORMANCE MEASURES FOR THE ANNUAL 1980 PERIOD.

	AES	AST	MOE	REL	RIV	RTM	SER	SIM	STA	TVA	UMA	OBS
I. Spatial Pattern Replication												
A. Magnitude												
1. Area of Significant Differences (%)*												
a. Northeast	10.5-	0.0	23.7-	1.9-	30.8-	22.2-	0.0	9.8+	29.7-	15.8-	27.1-	
b. North Central	1.8-	11.3+	2.9-	6.1	47.7-	50.0-	0.5-	6.4-	47.3-	0.9-	19.9-	
c. Southwest	2.2+	7.5+	0.9-	15.8+	55.3-	39.9-	0.0	0.4+	1.8-	5.3+	10.1-	
d. Southeast	47.4+	54.6+	1.3-	86.2+	21.1-	20.4-	3.3+	58.6+	0.0	30.9+	0.7+	
e. All 4 regions	10.4+	13.8+	7.6-	18.2+	42.9-	38.1-	0.7	13.3+	27.7-	9.7	17.7-	
2. Median (kg S/ha)	9.89	11.39	7.00	11.01	4.70	4.81	9.77	9.25	6.45	9.56	6.19	8.83
(p-value)	0.041	0.002	0.125	0.125	0.000	0.000	0.306	0.609	0.041	0.409	0.041	
3. Dispersion (kg S/ha)	14.93	23.40	10.08	53.76	1.73	1.17	4.46	18.04	11.55	20.23	10.23	10.67
(p-value)	0.181	0.017	0.439	0.000	0.000	0.000	0.010	0.078	0.415	0.042	0.455	
4. Bias-High (kg S/ha)	0.68	-3.36	2.03	-3.16	6.03	6.18	0.63	0.21	3.44	-0.60	3.01	
(p-value)	0.21	0.00	0.00	0.07	0.00	0.00	0.16	0.42	0.00	0.28	0.00	
5. Bias-Low (kg S/ha)	-1.38	-1.81	2.28	-2.60	2.88	2.24	-1.64	-0.72	1.89	0.04	2.25	
(p-value)	0.11	0.03	0.00	0.05	0.00	0.00	0.00	0.25	0.01	0.48	0.01	
6. Mean Squared Error(kg S/ha)^2	15.8	22.6	10.7	53.5	26.9	25.8	5.96	2.52	17.0	14.3	13.8	
B. Position												
1. Centroids (km)	119	76	58	134	43	75	46	114	119	77	61	
2. Major Axis (degrees)**	-0.4	-4.6	+7.8	+5.0	+3.9	+2.8	0.0	+6.9	+14.8	+8.8	-2.2	43.1
3. Shift (km)	397	330	220	467	311	348	330	397	330	311	348	
C. Shape												
1. Correlation-No Shift	0.00	0.39	0.75	-0.05	0.37	0.05	0.47	0.11	0.27	0.41	0.34	
2. Correlation-With Shift	0.64	0.66	0.82	0.61	0.79	0.73	0.83	0.62	0.66	0.76	0.77	
II. Maxima												
A. Location-Kriged (km)	202	273	0	202	0	55	273	178	138	69	168	
B. Magnitude-Kriged (%)**	0	+24	-5	+88	-48	-46	-6	+2	0	+3	-4	15.77
Magnitude-Site(%)**	0	+22	-4	+94	-47	-44	-5	+3	0	+4	-2	
III. Temporal Pattern Replication												
A. Seasonal Contribution (%)	.	.	.	.	.	.	.	.	.	.	.	
(Percent Difference)**	.	.	.	.	.	.	.	.	.	.	.	

*"+" and "-" denote the prevalence of over predictions and under predictions, respectively.

**Values listed for each model represent the value of the prediction minus the value of the observation.